HOW TO EVALUATE THE EFFECTIVENESS OF TEACHER TALK

Philip Glover

Süleyman Demirel University

pgaglover@gmail.com

Philip Glover has taught and trained teachers in Turkey, Cyprus, UK, Hungary and Serbia. He has worked in English language teaching as a manager and as a consultant with testing and in-service training projects.

HOW TO EVALUATE THE EFFECTIVENESS OF TEACHER TALK

Philip Glover

pgaglover@gmail.com

Abstract

Talk is a key component of teaching and learning and teacher talk is an important part of how a teacher teaches. This paper discusses principles and approaches for analysing teacher talk in classroom interaction in order to investigate its effectiveness. The paper presents four reasons for analysing teacher talk: the role of interaction in the management of learning, the analysable, interpretable and explainable nature of discourse, the value of analysing effective teaching and the existence of categories of teacher talk that support the investigation of effective teaching. These categories are based on a framework of Elicitations, Responses and Descriptions from Mercer (1995). The paper also looks at issues of reliability, validity and generalisability of findings. The paper concludes that a focus on categories of teacher talk can support academic research, for the supervision and evaluation of teachers by managers, as well as for self-development purposes by teachers themselves.

Keywords: classroom research, teacher talk, discourse analysis, elicitation*.*

1. Introduction: Classroom research, interaction and the effective management of learning

Classroom research investigates classroom events to produce insights into the learning process. Classroom research is about 'trying to understand what goes on in the classroom setting' and categorises classroom events as '*inputs to* the classroom (the syllabus, the teaching materials etc.)' and '*outputs from* the classroom' (Allwright & Bailey, 1991, p. 2). Inputs are present in the classroom in the form of teaching, lesson content or materials.

Interaction in teaching is a classroom input that has been widely debated and researched. Classroom interaction is encouraged because of its contribution to learning by facilitating lifelike or 'genuine' communication (van Lier, 1988). However, classroom interaction should not be viewed in the same way as other types of interaction. McDonough (2002, pp. 138-139), says there is a need to 'de-couple' natural language learning processes from classroom interaction in order to see the classroom as an environment which has its own language with its own purposes and characteristics. A specific characteristic of classroom language is that it has a pedagogical purpose, to facilitate learning. One way of looking at this pedagogical purpose is as part of the management of learning, which involves teachers and learners in a 'co-production' (Allwright & Bailey, 1991, p. 25). Teachers and learners contribute to the co-production in different ways, for example by talking.

Teachers make a key contribution to the management of learning through their talk. Mercer (1995, p. 1) describes communication 'in which one person helps another to develop their knowledge and understanding' as 'the guided construction of knowledge'. According to Mercer the role of the teacher can be crucial to the success of learning, and he notes classroom research that identifies ways for teachers to guide learners more effectively. This guiding is described as 'scaffolding', which is seen as 'a useful metaphor for the intellectual

involvement of a teacher with a learner's efforts during joint activity' (Mercer, 2000, p.159-170). Teaching that uses scaffolding thus makes an important contribution to learning.

A range of viewpoints have stressed the role of teaching and teachers in learning. The Russian psychologist Vygotsky, writing in the Thirties, noted the contribution to learning made by teachers (1962, 1978). Vygotsky saw the support and guidance that teachers provide in lessons through language as operating in a 'zone of proximal development' to facilitate learning. More recently advocates of teacher effectiveness have noted the importance of how teachers teach and have cited a range of language factors that contribute to learning such as explaining, questioning and interaction (Muijs & Reynolds, 2001).

Mercer (2000) also has a view on effectiveness, summarising research that found more effective teachers display three characteristics. Effective teachers firstly use questioning to guide the development of understanding and to test knowledge. Secondly, in addition to teaching the content of the subject, they use procedures to assist problem-solving and making sense of classroom experiences. Thirdly, they support learning as a social and as a communicative process.

Teachers' contribution to learning is achieved, at least in part, through teacher talk. Teachers use talk in lessons, and 'What is skilful and notable is the teacher's deployment of … language to get things done in the classroom' (Sinclair, 1982, p.10). In addition teacher talk has characteristics that are different from other types of talk: 'Teacher talk is different from doctor talk and preacher talk and all the many other kinds of talk… (teachers') *discourse* on the job is quite different' (Sinclair, 1982, p. 11). The study of teacher talk in lessons, then, is a way of looking at teachers' contribution to learning. One way of researching teacher talk is by analysing talk as discourse.

2. Discourse is analyzable, interpretable and explainable

Classroom research into how teachers teach looks at classroom events and teachers' actions in lessons. Analysing classroom talk, or classroom discourse, is one way of categorising teachers' actions and ascertaining their consequences. Classroom research has looked at features of classroom discourse for different purposes. Three features of discourse are particularly relevant to studies and evaluations of how teachers teach:

1 Discourse is action through talk that is analyzable.
2 Discourse varies between people and situations and is interpretable as showing purpose.
3 Discourse requires detailed analysis and can produce social explanations.

Feature 1- Discourse is action through talk that is analyzable.

People use discourse in the form of talk to get things done, for example teachers use language to get learning done. Talk is viewed as 'social action' (Mercer, 1995, p. 67), and discourse analysis is a means of describing and interpreting this action. Studying classroom discourse can help to explain 'what is being done in the discourse and how this is accomplished' (Wood & Kroger, 2000, p. 95). Studying discourse can help researchers and supervisors of teachers to understand the function of different utterances or 'what it is that people are doing in and with their talk and text' (Wood & Kroger, 2000:7). Looking at discourse can show how people use words to 'scaffold the performance of social activities (whether play or work or both)' (Gee, 1999, p. 1). In the context of the language lesson

teacher talk provides 'scaffolding' (Mercer, 2000, pp. 138-140) to support pedagogical purpose.

Discourse is therefore analyzable with patterns that can be revealed through study: 'Discourse can be regarded as sets of linguistic material that are coherent in organization and content and enable people to construct meaning in social contexts' (Cohen, Manion, & Morrison, 2000, pp. 245). Patterns in talk can support a better understanding of the talk and 'explore the organization of ordinary talk around everyday explanations and the social actions performed in them' (Cohen, et al., 2000, pp. 298). Patterns in the teacher talk, then, enable us to describe and interpret teachers' actions in the classroom.

Feature 2- Discourse varies between people and situations and is interpretable as showing purpose.

The second feature of discourse is that it is variable between individuals and situations Variations in discourse produce and reveal different viewpoints, as 'talk constructs different versions of the world and is oriented to different functions; variability is to be expected not only between persons, but within persons' (Wood & Kroger, 2000, p. 10). Variations in discourse can facilitate comparisons. Similarities and differences may be found between different teachers and between teachers in different situations.

Discourse is also interpretable as showing purpose and is not just a 'sequence of linguistic units' (Schiffrin, 1994, p. 416). Discourse involves a message, a medium of communication and intersubjectivity between the participants, a 'sharing of knowledge or experience', and participants 'use utterances to convey information and to lead each other towards an interpretation of meanings and intentions' (Schiffrin, 1994, p. 386). In turn the researcher also has to interpret meanings and intentions in order to describe and explain the pedagogical purpose of teacher talk. Discourse should be viewed as a system 'through which particular functions are realized' (Schiffrin, 1994, p. 32), so patterns in discourse can reveal not only the form of the talk but also the purpose or function.

The discourse is interpreted as evidence of the way of doing or teaching itself. Discourse is not viewed as evidence of teachers' beliefs or 'some internal event or attitude' (Wood & Kroger, 2000, p. 14). In this way a discourse approach can help to focus research on how teachers teach and avoid confusion with attitudes or cognition: 'the discourse itself is the educational reality' (Cohen, et al., 2000, p. 300 quoting Edwards, 1991).

Feature 3- Discourse requires detailed analysis and can produce social explanations

A third feature of discourse is that the researcher must be careful to deal thoroughly with the data and analysis in order to be convincing. A study requires detailed analysis of empirical data: 'discourse analysis requires a careful reading and interpretation of textual material' (Mason, 1996, p. 200). Qualitative research should be 'systematically and rigorously conducted', 'strategically conducted, yet flexible and contextual' and 'should involve critical self-scrutiny by the researcher, or active reflexivity' (Mason, 1996, p. 5). Any study of classroom discourse needs to be strategically oriented, paying attention to established views of teaching, but the study also needs to be flexible, taking context into account. Discourse analysis should produce social explanations: '… meaningful elements in a complex… social world', '… based on methods of data generation which are flexible and sensitive to the social context' and '… based on methods of analysis and explanation building

which involve understandings of complexity, detail and context' and findings '… should produce social explanations which are generalisable in some way, or which have a wider resonance' (Mason, 1996, p. 6). Discourse analysis can reveal positive or negative effects of teacher talk. Even a small-scale study of classroom discourse can show action that is repeated by many teachers and students.

3. Effective teaching

The school and teacher effectiveness literature is based on international research in Britain, the United States, the Netherlands and a number of other countries (Campbell, Kyriades, Muijs, & Robinson, 2003; Creemers & Reynolds, 1996; HayMcBer, 2000; Muijs & Reynolds, 2001; Preedy, 1993; Reynolds, 1999; Reynolds, Davies, & Phillips, 1989; Reynolds & Muijs, 2000a, 2000b, 2000c, 2000d; Sammons, 1999; Sammons, Hillman, & Mortimore, 1995; Tabberer, 1994; Teddlie, 2000). The research notes the importance of teaching behaviours, identifying 'concentration on teaching and learning' and 'purposeful teaching' as factors producing effective schools (Sammons, et al., 1995, pp. 13-15). 'Interactive teaching' makes an important contribution to teaching and learning and involves such features as 'direct instruction', teachers listening and responding to students and providing clear instructions and explanations (Muijs & Reynolds, 2001, pp. 17-26).

Teacher effectiveness research states that interactive teaching (Muijs & Reynolds, 2005) is more effective than teaching with an emphasis on individual work, although individual work in combination with interaction also supports learning. Whole-class teaching is thought to be effective because it 'allows the teacher to make more contacts with each individual pupil' and allows teachers to monitor and respond to students more quickly (Muijs & Reynolds, 2001, pp. 5). This enables teachers to adapt classroom procedures to meet student needs and keep students on task. Appropriate and varied questioning, probing for knowledge and frequent feedback are factors in interactive teaching (Reynolds & Muijs, 2000d). Nine factors contribute to effective interactive teaching: when and how often to use questioning, eliciting a pupil response, the cognitive level of questions, open and closed questions, process and product questions, what to do when a student answers correctly, what to do when a student answers correctly but hesitantly, what to do when a student answers incorrectly, what to do when a student answers partially correctly, and prompting (Muijs & Reynolds, 2005). A number of indicators contribute to effective interaction. Questioning should take up a large part of the lesson and involve 'guided practice'. Eliciting should bring out a response from students in a positive, non-evaluative atmosphere. The cognitive level or difficulty of questions should require thought on the part of students and should be challenging but within the capacity of students to answer. Effective questioning involves asking more open questions and more process questions. Effective interaction acknowledges correct responses, responds with positive feedback to correct hesitant student utterances and provides supportive feedback to incorrect student utterances (Muijs & Reynolds, 2005). Effective teaching therefore involves prompts using cues, reminders or references to previous learning.

In addition to interaction through questioning, the teacher effectiveness literature also comments on teacher talk that passes on information to students. Muijs and Reynolds (2005) cite types of information-giving talk: directing, instructing, demonstrating, explaining and illustrating, consolidating, evaluating and summarising. Effective directing involves the teacher sharing the aims of a lesson and clarifying points for particular attention. Effective instructing involves giving students information in a clear, comprehensible and structured way, and effective demonstrating involves showing or providing models. Explaining should be accurate, at an appropriate moment in the lesson and should help the students to make connections with past experiences. Consolidating should reinforce and develop points and

encourage student reflection. Evaluating identifies student errors and summarising involves reviewing what has been done in a lesson (Muijs & Reynolds, 2005). Talk that presents information may also use questioning or prompting to support or guide learners' understanding.

There are many common elements between the teacher effectiveness literature and Mercer's view of the 'guided construction of learning' (Mercer, 1995). Both are concerned with improving teaching. Both note the importance of the contribution of teachers to learning. Both note the role that discourse, especially interactive discourse, plays in the teaching and learning process. Analysis is not just a question of describing classroom discourse but also evaluating how well it achieves its purpose, and how well the 'ways of guiding the construction of knowledge… seem suited to the kinds of learning they are supposed to encourage' (Mercer, 1995, p. 41). These common elements provide a foundation for analysing classroom discourse and identifying the effectiveness of teacher talk.

4. Discourse analysis in language classrooms

Discourse analysis has been used extensively to analyse discourse in language classrooms. Studies have used a range of different methodologies to produce insights into classroom interaction. These studies show the potential benefits of applying discourse analysis in education and applied linguistics. The studies also indicate the importance of an analytical framework to fit the purpose and context of the research.

Sinclair (1982) studied classroom discourse and identified a pattern consisting of 'Initiation' by the teacher, 'Response' by students and 'Follow-up' by the teacher again. This is often referred to as the 'IRF' pattern. One of the purposes of Sinclair's study was to help teachers 'monitor their own performance' (Sinclair, 1982, p. 5). The 'IRF' pattern was subsequently used by studies to analyse and comment on many aspects of language teaching (Cadorath & Harris, 1998; Cullen, 1998; Dinsmore, 1985; Duff, 2002; Hall, 1998; Kumaravadevilu, 1993; Nunan, 1987; Ohta, 1999; Thornbury, 1996).

Several of these studies contrasted patterns in classroom talk with other types of communication (Kumaravadevilu, 1993; Nunan, 1987). They pointed out that the structure of classroom talk was dominated by one of the participants, the teacher, and this domination denied the other participants, the students, the opportunity to make language choices. This led to claims that 'in communicative classes interaction may not be very communicative at all' (Nunan, 1987, p.144). Others suggested that even teachers who support communicative teaching 'can fail to create opportunities for genuine interaction' (Kumaravadevilu, 1993, p. 13). The presence of the 'IRF' pattern has therefore been used to support a view that classroom interaction should replicate real-life communication in order to facilitate learning.

Alternative views of classroom discourse, however, see talk in lessons in a different light, with the function of guiding and supporting learning. According to this view it is more important to understand how 'IRF' patterns 'relate to the core institutional goal rather than dismissing them as undesirable or not genuine' (Seedhouse 1996, p. 22).

Discourse analysis has been used to produce a variety of insights into classroom talk and how teachers teach. For example, a study looked at issues of identity, respect and language socialisation in a school in Canada, investigating 'the co-construction of knowledge, identity and difference in/through classroom discussion in high schools' (Duff, 2002, p. 295). The study focused on two lessons to provide insights into relations between students and into teacher's unsuccessful attempts to 'make cultural connections based on their (the students') own backgrounds, cultures and experiences' (Duff, 2002, p. 310). Mori (2002, p. 325) looked at 'the task designer's intentions and 'notions of 'authenticity' and 'naturalness' and student

talk during a language learning task. Another study used systemic functional grammar to analyse the 'socialisation of students to science discourse' (Young & Nguyen, 2002, p. 348). This study compares language used by a teacher in a lesson with language in a text book. These studies point to two specific factors. Studies need to link methodology and the 'needs of educators' (Mori 2002, p. 341) in order to produce useful results and the approach used needs to fits the aims of the study and the context. The framework and categories described in this paper are introduced in Glover (2006) and the results were reported in Glover (2014).

5. Analysing teacher talk: Elicitation, response, description

There are several requirements for a framework for the analysis of teacher talk. In order to investigate the effectiveness of teacher talk the framework needs to:

○Consider pedagogical purpose focusing on the contribution of teachers to learning.

○Investigate patterns that make a positive contribution to learning.

○Fit the aims and context of the study.

A framework based on teachers' guidance strategies from Mercer (1995) provides a match with these requirements. Mercer's framework is based on the concept of the 'guided construction of knowledge' which he puts at the heart of the combined process of teaching and learning that makes up education.

Mercer looks at pedagogical purpose, especially the teacher's contribution to learning. He notes that teachers make an important contribution through guidance strategies (Mercer, 1995). These strategies are 'intentional, goal-directed ways of talking… which reflect the constraints of the institutional setting'. Teachers 'use talk' to 'guide learning' with three techniques: eliciting knowledge from students, responding to what the students say and describing shared classroom experiences (Mercer, 1995, p. 25-26). These techniques form the three main categories of analysis: Elicitation, Response and Description.

Mercer divides eliciting, responding and describing into sub-categories that relate to different types of teacher talk in the classroom. Mercer identifies two types of Elicitation, Direct and Cued (Mercer, 1995, p. 26-32). Direct Elicitation uses questions to bring out knowledge from students. Cued Elicitations encourage student talk by providing a clue or prompt that leads to a student utterance, for example a teacher eliciting the word 'pulse' by saying 'you can feel it here'(Mercer, 1995, p. 27).

The Response category refers to teacher responses, not student responses, unlike the IRF pattern which puts student responses as the second step in the sequence. Teachers use Confirmation, Rejection, Repetitions, Reformulations or Elaborations (Mercer, 1995, p.32-33). In Confirmations and Rejections a teacher may say 'yes' or 'no', 'right' or 'wrong'. Repetitions draw the attention of the class to a remark thought by the teacher to be significant. Reformulations offer a revised version of what has been said. In elaborations a teacher extends a student statement.

In the Describe category teachers use 'we' statements, Literal Recaps and Reconstructive Recaps (Mercer, 1995, p. 33-41). 'We' statements talk about a past experience that is relevant to the present, for example when a teacher reminds the students about something that happened in the previous lesson or reviews what has happened in the current lesson. Recaps review aspects of shared knowledge, for example, through reminders about previous lessons or drawing conclusions after a discussion or activity. Literal Recaps repeat the shared knowledge and Reconstructive Recaps add further interpretation.

These guidance strategies contribute to learning. They provide a means of both identifying patterns in talk, and evaluating the contribution the talk makes to learning. Mercer recognises the presence of the 'IRF' pattern within these categories, describing it as 'the 'classic' format of teacher-pupil interactions' (Mercer 1995, p. 38). He recognises the constraints that the pattern puts on pupil contributions to lessons, but he also points out that teachers do more than just initiate and follow up student utterances.

The framework using Elicitation, Response and Description categories can enable the researcher to consider pedagogical purpose and focus on the contribution of the teacher to classroom events. The framework relates to any general classroom context, so it may require adaptation to fit a language learning context in a specific place. The framework can help to identify patterns in teacher talk.

Teacher effectiveness measured by analysis and interpretation of teacher talk should not be a simple matter of counting easily identified indicators. Effectiveness cannot be brought down simply to the length of teacher talking time or other numbers, and teacher effectiveness research has been criticised for becoming 'a reductionist mode of research… replacing sociological analysis with increasingly complex statistics' (Wrigley, 2003, p .89). Teacher effectiveness researchers themselves warn that research requires careful reflection by practitioners and: 'a rather large "health warning" attached to any mandated set of teaching behaviours that may emanate from national policy makers' (Muijs & Reynolds, 2005, p . 265). Instead, evaluating the effectiveness of teacher talk needs to investigate how the different categories and sub-categories of the talk combine with each other.

Two key aspects of classroom discourse that feature in teacher effectiveness literature are interaction and providing information. Interaction involves questions, prompts, evaluation and correction by the teacher. Giving students information involves explanations and instructions. These aspects of classroom discourse match the Elicitation, Response and Description categories and are summarised in Tables 1 and 2:

Table 1. *Mercer's elicitation and response categories, effective questioning factors and indicators of effective teaching*

Mercer's Categories	Effective questioning factors	Indicators of effective teaching
Direct and *Cued Elicitation*s	When and how often to use questioning	Questioning takes up a large part of the lesson
Direct and *Cued Elicitation*s	Eliciting a pupil response	Questioning brings out responses from students
*Direct Elicitation*s	The cognitive level of questions	There are more questions that require the students to think, answers are not easy
*Direct Elicitation*s	Open and closed questions	More open questions are asked in effective lessons
*Direct Elicitation*s	Process and product questions	More questions about processes are asked
*Cued Elicitation*s	Prompting	Students are guided through cues, reminders or references
Confirmation *Responses*	What to do when a pupil answers correctly	Correct answers are acknowledged
Confirmation *Responses*	What to do when a pupil answers correctly but hesitantly	Uncertainty is clarified
Repetition *Responses*	What to do when a pupil answers incorrectly	Incorrect answers are corrected or a correct answers is prompted

Table 1 summarises the match between Elicitation and Response categories and teacher effectiveness criteria. The sub-categories of Direct and Cued Elicitations connect with when and how often to use questioning and eliciting a pupil response. Direct Elicitations connect

with the cognitive level of questions, for example through open and closed questions and process and product questions. Cued Elicitations connect with prompting. Confirmation Responses relate to what a teacher does when a pupil answers correctly or correctly but hesitantly, and Rejection or Repetition Responses relate to what a teacher does when a pupil answers incorrectly.

Table 2 summarises the match between Mercer's Description category and the teacher effectiveness categories of directing, instructing, demonstrating, explaining and illustrating, consolidating, evaluating and summarising.

Table 2. *The description category, teacher effectiveness categories and indicators of effective teaching*

Category	Teacher effectiveness categories	Indicators of effective teaching
Descriptions (recaps or 'we statements')	Directing	Teacher shares the aims of a lesson or clarifies points for particular attention.
	Instructing	Teacher gives students information in a clear, structured way.
	Demonstrating	Teacher shows or provides models
	Explaining and illustrating	Teacher explanations are accurate, at an appropriate moment in the lesson and help the students to make connections with past experiences.
	Consolidating	Teacher reinforces and develops points and encourages student reflection.
	Evaluating	Teacher identifies student errors as positive teaching points as well as talking about student justifications of their answers.
	Summarising	Teacher reviews what has been done in a lesson

There are several different ways of gathering data for classroom research involving discourse analysis. Audio or video recordings of lessons can be used, either by listening to the recording and taking notes or by transcribing the talk (for example van Lier, 1988; Lyle, 2002; Mori, 2002). Studies have also used observation schedules such as COLT or FOCUS as described in Allwright and Bailey (1991). Observation schedules look at classroom discourse by using predetermined categories. Such schedules are useful because they provide a clear focus for observers and enable different observers to focus on the selected features of classroom events. Classroom events are complex and much happens at the same time in classrooms so the schedule helps observers focus on specific aspects.

Observation schedules have some drawbacks, however. Compared to transcriptions of audio or video recordings they provide a less direct and less flexible version of classroom events. Schedules may not anticipate all relevant classroom features in advance, so a given schedule may not fit the purposes of a given study. For example, in one study 'several significant activities were not specifically identified by either COLT or the UCLES instrument' (Hayes & Read, 2004, p .103). Schedules produce a simplified version of an observer's interpretation of what was said at a given moment: 'the thirst to operationalise concepts and constructs can easily lead researchers to provide simple indicators of complex concepts' (Cohen, et al., 2000, p. 310). Observation schedule categories and observer's interpretation become the data rather than the actual words used by the teacher.

The direct and flexible nature of transcribed recordings make them appropriate to studies of classroom discourse and address sampling issues. The audio recording allows transcribing and analysing all the teacher talk that occurs in a given lesson. This is a kind of 'theoretical' or 'purposive' sampling, which 'means selecting groups or categories to study on the basis of their relevance to your research questions, your theoretical position and analytical framework, … and … the explanation or account which you are developing' (Mason, 1996, p .93-94).

Transcribed teacher talk provides a direct representation of how teachers teach. The talk is classroom action, one of the ways in which teachers perform the functions of teaching or guiding the construction of knowledge in learners. The talk shows whether and how the teacher might influence the learning process using an analytical framework based on Elicitations, Responses and Descriptions.

Data from field notes, interviews and background data can also support an analysis of teacher talk. Transcribed talk can provide a detailed and even a complete picture of teacher talk in a lesson, but it cannot capture everything that happens in a lesson. The use of multiple data sources is therefore recommended as a means of providing a fuller picture of classroom events (Allwright & Bailey, 1991).

6. Discussion

In the language classroom specific categories and sub-categories of teacher talk provide a way to evaluate the effectiveness of the teaching. Teacher talk data may be in the form of recordings of lessons, observation schedules or field notes. Having ascertained which types of teacher talk are present in the data or lesson, conclusions then need to be reached about the effectiveness of the talk. This can be done by a combination of quantitative and qualitative methods. Quantities to measure are the absence or overuse of specific categories and sub-categories. Qualities to identify are the sequences or combinations of the categories and sub-categories as well as other questions such as the comprehensibility, level of challenge or interactive nature of the talk. Just looking at numbers of occurrences can be deceptive, because what really makes talk effective are the qualities of that talk. A combination of both quantitative and qualitative methods is therefore likely to be beneficial to a study.

Mercer's categories and sub-categories are summarised in Table 3. Investigating these categories and sub-categories allows a researcher to look in close detail at both the quantities and qualities present in the discourse to see how a teacher teaches and to see the effectiveness of the teaching.

Table 3. *Mercer's categories and sub-categories*

Category	Sub-category
Elicitations	Direct
	Cued
Responses	Confirmations
	Rejections
	Repetitions
	Elaborations
	Reformulations
Descriptions	we statements
	Literal Recaps
	Reconstructive Recaps

In terms of quantities, the presence or absence or the number of occurrences of each category or sub-category can give an impression of how a teacher teaches. For example, if a

lesson consists mainly of Descriptions, with few or no Elicitations and Responses, then the effectiveness of the lesson is likely to be low. If the teacher spends most of the lesson giving information to the students without interaction there is no evidence in the discourse that the students are able to understand or engage with what the teacher is saying. Similarly, if the talk is heavily dependent on one type of Elicitation or Response, for example yes-no questions, which belong only to the sub-categories of Confirmations or Rejections, then the level of cognitive challenge is likely to be low.

As for the qualities of the talk, the researcher can consider how the different categories and sub-categories fit together into sequences or combinations. The way a teacher follows up student utterances makes an important contribution to the effectiveness of the discourse. For example, after an elicitation, if a teacher responds to a student's incorrect utterance simply by stating what is correct, that is likely to be less effective than a series of further elicitations that guide the student to the correct answer by encouraging self- or peer-correction.

Cognitive challenge can be investigated by looking at different types of elicitations. Elicitations can be described as direct or cued. Direct elicitations are questions, of which there may be different types. The level of cognitive and linguistic challenge of questions can vary according to the type of questions. Levels of difficulty start with simple yes-no questions, questions offering options are slightly more challenging, and open or wh-questions can be more concrete or more abstract. Questions may be about processes or concepts, for example in language teaching concept questions are a technique that can be used in language classrooms to clarify structures or the meaning of vocabulary. Cued elicitations are prompts that bring talk out of learners, for example by means of the teacher starting an utterance for learners to complete, requests for the learner to say more, wait time or back-channel encouragement.

Teachers can use a succession of Elicitations and Responses to encourage learners to think about the language they are using and to coax talk out of the students. When teachers use Elicitations they work with students to build knowledge together. this 'embodies an educational process in which the pupils are neither being drawn out of themselves… nor simply being taught directly in the 'transmission' sense' because when teachers use *Cued Elicitations* to bring talk out of students, students 'are being inculcated into what becomes for them a shared discourse with the teacher (discourse in the broadest sense, including concepts and terminology as well as dialogue)'. This is 'a communicative process of substantial intrinsic interest' where 'pupils' knowledge is aided and 'scaffolded' by the teacher's questions, clues and prompts to achieve insights that the pupils themselves initially seem incapable of' and as a result 'effective scaffolding reduces the learner's scope for failure in the task, while encouraging their efforts to advance' (Edwards & Mercer, 198,p. 143). In these ways, how a teacher combines different types of Elicitations and Responses can be seen to contribute to effective teaching.

The researcher or observer should also consider other types of talk such as the Descriptions. In an English language class Descriptions are likely to involve reviews of learning from previous lessons or the current lesson, instructions concerning a task that is to be done, or explanations of language that is being studied such as grammar or vocabulary. Descriptions are likely to be less effective if they are simply a monologue where the teacher tries to pass information or knowledge to the students. Descriptions are likely to be more effective if they are combined with Elicitations and Responses which check comprehension, for example asking students to apply a rule, put a word into a sentence or repeat back instructions for a task.

Indicators of effective teaching that may be found in Elicitations, Responses and Descriptions are as follows. Questioning takes up a large part of the lesson and brings out responses from students through the use of some questions that require the students to think such as more open questions. In effective lessons more questions about processes are asked, correct answers are acknowledged and uncertainty is clarified. Incorrect answers are corrected or a correct answer is prompted or guided through cues, reminders or references. A combination of different types of Elicitations and Responses are therefore needed to achieve effectiveness. The indicators of effective teaching that may be found in Descriptions are that at the start the teacher shares the aims of a lesson or clarifies points for particular attention, gives students information in a clear, structured way, and shows or provides models, teacher explanations are accurate, at an appropriate moment in the lesson and help the students to make connections with past experiences, the teacher reinforces and develops points and encourages student reflection and identifies student errors as positive teaching points as well as talking about student justifications of their answers. At the end of a lesson the teacher reviews what has been done in a lesson. Consideration of all these criteria build up a picture of the teacher talk in terms of its effectiveness.

7. Reliability, validity and generalizability

Data analysis will involve a 'detailed and repeated reading of the discourse against the discourse-analytic perspective' (Wood & Kroger, 2000,p. 95). Analysis can be supported by computer software for qualitative data analysis. Measures of difference will be overall patterns in the discourse as revealed by the analytical framework, qualities of the categories and sub-categories in the framework, the presence or absence of certain types of categories or sub-categories and the frequency of occurrences.

Detailed reading and analysis of the data follows a recursive process of 'noting', 'collecting', 'thinking' and 'comparing' (Seidel, 1998). coding the data according to functions and according to the categories of the analytical framework. Looking at each utterance in turn involves deciding whether words or groups of words are *Elicitations*, *Responses* or *Descriptions*, and if so what type of each category they are. An *Elicitation* could be *Direct* or *Cued*, for example.

The next stage in the process involves collecting together the examples of each sub-category and sorting them into groups that appeared to be similar in terms of their function. Whilst doing this the groups that are forming, how they form into patterns and how they relate to the overall framework must be considered. At the same time the researcher checks and improves the original coding.

Analysis can be supported by a suitable computer program for qualitative data analysis. Computer assisted qualitative data analysis has advantages such as speed and rigour, as well as the disadvantage that the tool itself rather than methodological principles may influence the study (Seale, 2000). Two popular computer software packages are ATLAS-ti and NUD*IST. ATLAS-ti has been reviewed as more user-friendly, simpler and easier to use, whereas NUD*IST offers more sophisticated analysis possibilities (Barry, 1998; Lewis, 1998). Atlas-ti's strengths lie in its 'interconnectedness and creative interface", whereas NUD*IST is stronger on 'sequential structure", 'project management' and 'sophisticated searching' (Barry, 1998, p .12.1).

Evidential links between the data and the effectiveness framework are crucial for the identification of effective teaching. The first means of establishing evidential links is in the reliability and accuracy of the data and the coding. The second means of establishing evidential links lies in the validity of the analysis. The third means of establishing evidential

links lies in the generalisability of explanations. These three elements contribute to producing analyses and explanations which are convincing (Mason, 1996). Convincing claims depend on attention to reliability, validity and generalisability, and the 'construction of explanations needs to be done with rigour, with care, and with a great deal of intellectual and strategic thinking' (Mason, 1996, p .162).

For reliability the effectiveness of teacher talk can be measured by the quantities and qualities of the discourse categories. Differences in quantities may occur in the number of occurrences of the categories and sub-categories in the analytical framework. There must be a reliable means of coding judgements as to what utterances belong to each category and sub-category. There also needs to be a coherent and reliable way of counting the categories: the amount of time spent on each category, the number of incidences of each category or the number of words used.

Differences in qualities of the teacher talk may also occur. How the teacher applies the categories may be different in terms of function, pedagogical purpose, words used or content. Coding must be consistent and accurate. Grouping utterances in each category must be based on qualities that are relevant to the study.

The reliability of the method also depends on the validity of data generation (Mason, 1996, p .145-146). Data in the form of transcribed teacher talk must provide a means of analysing how teachers teach (see Chapter 2.1 and 2.2 above). During analysis the data must fit the framework chosen. Then there needs to be evidential links between the data and the categories. The quantities and qualities of *Elicitations*, *Responses* and *Descriptions* represent judgements of how the teacher is teaching and how that teaching relates to the examination (Mason, 1996, p. 146).

The use of multiple measures, viewpoints and analyses will add to the accumulation of judgements provided. However, Mason warns against expecting multiple measures to corroborate each other in providing more evidence of the same phenomenon. Different analyses imply different explanations, not more evidence for the same explanation (Mason, 1996, p .149). The value of multiple analyses will be in the additional evidence they present rather than in strengthening previous explanations.

The validity of interpretation is shown by explaining and justifying the logic of methodological choices and by describing the route by which interpretation is reached, 'continually and assiduously charting and justifying the steps through which (my) interpretations are made' (Mason, 1996, p. 150).

Finally the analysis must demonstrate that the explanations have 'plausibility' and 'fruitfulness' (Wood & Kroger, 2000, pp. 174-175) in terms of their generalisability and value (Mason, 1996, pp. 152-155). This will involve consideration of three points:

1. How the experiences of the teachers are relevant to other teachers' experiences. Analysis must be detailed and holistic in order to show that what happens in the selected lessons can happen in other classrooms.

2. Analysis must show 'the wider resonance or generalisability of (my) explanations' 'based on the rigour of my analysis' (Mason, 1996, p. 154) by showing that explanations can produce benefits for teachers and that there are no adverse consequences for the teachers in the study.

3. The use of 'aggregation, numbers and counting in a meaningful fashion' (Mason, 1996, p .155) must address issues of sampling and significance.

8. Conclusions

This paper has argued that a focus on the categories of Elicitation, Response and Description in teacher talk can help to develop an understanding of how classroom interaction can be more effective. The approach outlined can be used primarily by academic research for investigating how teachers teach and its effectiveness for learning. The approach can also be adapted and used for the supervision and evaluation of teachers by managers. The categories can support teacher supervision by establishing criteria that can be agreed upon by the supervisor and the teacher. The categories also provide supervisors and teachers with common language that can make feedback sessions quicker and more productive because they facilitate discussions as in Walsh (2002, 2006). Teachers' self-development can also be supported by having such language to assist teachers in reflecting on what happens in their own lessons.

References

Allwright, D., & Bailey, K., M (1991). *Focus on the language classroom*. Cambridge: Cambridge University Press.

Barry, C. A. (1998). Choosing qualitative data analysis software: Atlas/ti and nudist compared. *Sociological Research Online, 3*(3), <http://www.socresonline.org.uk/socresonline/3/3/4.html>.

Cadorath, J., & Harris, S. (1998). Unplanned classroom language and teacher training. *ELT Journal, 52*(3), 188- 196.

Campbell, R. J., Kyriades, L., Muijs, R. D., & Robinson, W. (2003). Differential teacher effectiveness: Towards a model for research and teacher appraisal. *Oxford Review of Education, 29*(3), 347- 362.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education*. London: Routledge.

Creemers, B. P. M., & Reynolds, D. (1996). Issues and implications of international effectiveness research. *International Journal of Educational Research, 25*(3), 257-266.

Cullen, R. (1998). Teacher talk and the classroom context. *ELT Journal, 52*(3), 179-187.

Dinsmore, D. (1985). Waiting for Godot in the EFL classroom. *ELT Journal, 39*(4), 225-234.

Duff, P. A. (2002). The discursive co-construction of knowledge, identity, and difference: a ethnography of communication in the high school mainstream. *Applied Linguistics, 23*(3), 289-322.

Edwards, D., & Mercer, N. (1987). *Common knowledge: The development of understanding in the classroom*. Cambridge: Cambridge University Press.

Gee, J. P. (1999). *An Introduction to discourse analysis theory and method*. London: Routledge.

Glover, P. (2006). *Examination influence on how teachers teach: A study of teacher talk* (Unpublished PhD Thesis). Lancester.

Glover, P. (2014). Do language examinations influence how teachers teach? *International Online Journal of Education and Teaching, 1*(3), 197-214.

Hall, J. K. (1998). Differential teacher attention to student utterances: the construction of different opportunities for learning in the IRF. *Linguistics and education, 9*(3), 287-311.

Hayes, B., & Read, J. (2004). IELTS test preparation in New Zealand: Preparing students for the IELTS academic module. In L. Cheng, Y. Watanabe & A. with Curtis (Eds.), (pp. 3- 18). New Jersey: Laurence Earle Brown Associates.

HayMcBer (2000). *Research into teacher effectiveness: A model of teacher effectiveness*. London: Department for Education and Employment.

Kumaravadevilu, B. (1993). Maximizing learning potential in the communicative classroom. *ELTJ, 47*(1), 12-21.

Lewis, B. R. (1998). ATLAS/ti and NUD·IST: A Comparative review of two leading qualitative data analysis packages. *Cultural Anthropology Methods, 10*(3), 41-47.

Lier, L. v. (1988). *The classroom and the language learner:ethnography and second-language classroom research*. Harlow: Longman.

Lyle, S. (2002). Talking to Learn: The voices of children, aged 9–11, engaged in Role-play. *Language and Education, 16*(4), 303-317.

Mason, J. (1996). *Qualitative researching*. London: Sage.

McDonough, S. (2002). *Applied linguistics in language education*. London: Arnold.

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners*. Clevedon: Multilingual Matters Ltd.

Mercer, N. (2000). *Words and Minds: How we use language to think together*. London: Routledge.

Mori, J. (2002). Task design, plan, and development of Talk-in-interaction: An analysis of a small group activity in a Japanese Language Classroom. *Applied Linguistics, 23*(3), 289-322.

Muijs, D., & Reynolds, D. (2001). *Effective Teaching: Evidence and practice*. London: Sage.

Muijs, D., & Reynolds, D. (2005). *Effective Teaching: evidence and practice* (2nd ed.). London: Sage.

Nunan, D. (1987). Communicative language teaching: making it work. *ELT Journal, 41*(2), 136-145.

Ohta, A. S. (1999). Interactional routines and the socialisation of interactional style in adult learners of Japanese. *Journal of pragmatics, 31*, 1493-1512.

Preedy, M. (1993). *Managing the effective school*. London: Open University.

Reynolds, D. (1999). *Better teachers, better schools*: Pearson Publishing.

Reynolds, D., Davies, R., & Phillips, D. (1989). The Cardiff programme: an effective school improvement project based upon school effectiveness research. *International Journal of Educational Research, 13*(7), 801- 814.

Reynolds, D., & Muijs, D. (2000a). *Advanced teaching methods* (pp. 8). Pearson Publishing.

Reynolds, D., & Muijs, D. (2000b). *Effective teaching: British research* (pp. 6). Pearson Publishing.

Reynolds, D., & Muijs, D. (2000c). *Effective teaching: international research* (pp. 6). Pearson Publishing.

Reynolds, D., & Muijs, D. (2000d). *It's the classroom stupid* (pp. 6). Pearson Publishing.

Sammons, P. (1999). *School effectiveness*. Lisse, Exton, PA: Swets & Zeitlinger.

Sammons, P., Hillman, J., & Mortimore, P. (1995). *Key characteristics of effective schools*. London: Office for Standards in Education.

Schiffrin, D. (1994). *Approaches to discourse*. Oxford: Blackwell.

Seale, C. (2000). Using computers to analyse qualitative data. In D. Silverman (Ed.), *Doing qualitative research; a practical handbook* (pp. 154-174). London: Sage.

Seidel, J. V. (1998). *Qualitative data analysis* (pp. 15). Retrieved from www.qualisresearch.com

Sinclair, J. (1982). *The structure of teacher talk*. Birmingham: University of Birmingham.

Tabberer, R. (1994). *School and teacher effectiveness*. Slough: National Foundation for Educational Research.

Teddlie, C., & R., David (2000). *The international handbook of school effectiveness research*. London and New York: Falmer Press.

Thornbury, S. (1996). Teachers research teacher talk. *ELT Journal, 50*(4), 279-289.

Vygotsky, L. S. (1962). *Thought and language*. Cambridge: M.I.T. Press, Massachusetts Institute of Technology.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.

Walsh, S. (2002). Construction or obstruction: Teacher talk and learner involvement in the EFL classroom. *Language teaching research, 6*(1), 3-23.

Walsh, S. (2006). Talking the talk of the TESOL classroom. *ELT J, 60*(2), 133-141.

Wood, L. a., & Kroger, R. O. (2000). *Doing discourse analysis Methods for studying action in talk and text*. London: Sage.

Wrigley, T. (2003). Is 'School Effectiveness' anti-democratic? *British Journal of Educational Studies, 51*(2), 89-112.

Young, R. F., & Nguyen, H. T. (2002). Modes of meaning in high school science. *Applied Linguistics, 23*(3), 348-37*2*

DIGITAL LITERACIES AND INTERACTIVE MULTIMEDIA-ENHANCED TOOLS FOR LANGUAGE TEACHING AND LEARNING

Hülya Mısır

Ufuk University

hulyamsr@gmail.com

Hülya Mısır is a research assistant at Ufuk University, English Language Teaching Department holding an M.A. in English Language Teaching at Hacettepe University. Her interests include digital literacies, online language learning environments, language MOOCs, individual differences, autonomy, and multilingualism.

DIGITAL LITERACIES AND INTERACTIVE MULTIMEDIA-ENHANCED TOOLS FOR LANGUAGE TEACHING AND LEARNING

Hülya Mısır

hulyamsr@gmail.com

Abstract

The primary aim of the paper is to introduce digital literacies and the utilization of interactive and multimedia-enhanced tools for language learning purposes in and out of institutional settings. The topics covered in this paper include what digital literacy means in language education contexts and utilization of social media, online gaming, tagging, picture, voice, and video editing tools, mixing tools, interactive HTML5 contents, translation tools, search engines and (visual) thesauruses to evaluate, contribute, and create the information rather than being a mere consumer. Bearing in mind the nature of interaction and collaboration of the digital tools, suggested ways to integrate these tools to teaching and learning of languages are also presented. The paper also identifies common concerns and challenges in regard to motivation to use digital tools, judging the validity of a source, prohibitions due to country policies, plagiarism, filtering in search engines, and insensitivity with online language, which hinders the effective use of digital tools.

Keywords: digital inclusion, digital literacy, digital tools, e-skills, new literacies

1. Introduction

There is a rapid pace of change in terms of information technology. Digital tools, thus, become more and more widespread globally in many areas of our lives. In the case of language teaching, teachers now have to take changing needs of learners of this new digital era and related capabilities into consideration. As Godwin (2015) states, capability in a full range of digital literacies is a key for this process and vital today in education, personal life, and work environments. Most importantly since the medium language is English, to take full advantage of the new possibilities and opportunities that information technology offers, English language learners have to adapt the latest literacy to their education (in-class learning) and personal life (out-of-class learning).

Competences regarding digital literacy can be taught in a formal class environment; however, the ultimate success requires active participation in web platforms and practice to be a natural part of digital environments. Learners can be trained to hybridize technology-mediated input and competences of utilizing digital tools to reach an answer or conclusion, participate in generating discussions or ideas in social platforms and collaborative team work, and contribute to the ultimate output, and circulation of knowledge. Therefore, teachers highlight that the process includes both formal and informal way of learning, and learners need to explore more to be a better participator and contributor. Godwin (2015) advocates the point stating that teachers help students become lifelong learners and integrating language learning into everyday online practices is essential. As learners manage to take place in interactive environments, they develop autonomy by taking the responsibility of their own language learning. The major role of teachers in this regard is to encourage learners to acquire skills of choosing and processing the knowledge. Therefore, digital technology not only offers "more options for teachers and learners, but it directly affects the roles of both

teachers and learners" (Stockwell, 2015, p. 362). As teachers adapt more learner-oriented/centered approach, this can motivate learners to develop learning strategies and promote autonomous language learning.

In SLL or FLL contexts, the traditional view is that the use of techie tools helps learners improve language skills (Tour, 2012). However, today digital technology necessitates the engagement between the tools and learners in a different way. It is beyond a bilateral relationship and includes interaction with other users of the same tools/platforms. The learners who recognize the interactive nature of digital language tools involve in the participatory culture of digital/online platforms by being a creator rather than a mere consumer of passive media materials. Accordingly, they become an integrated part of a globalized context bringing new social construct with respect to networking, collaboration, interaction and motivation that can be truly managed with digital competence.

In the recent past, to promote communicative language learning in class, bringing technology into classroom meant passive media materials such as movies, videos, prints of newspapers or magazines etc. to language teachers. Calling them passive is not unfair by all means since they did not aim at developing "strategic competencies of digital literacy encapsulating multimedia, collaborative communication, agentive participation, and multitasking competencies" (Lotherington & Ronda, 2014). These passive in-class materials, though they are authentic and real-world materials, taught *teachers* how to make use of the materials offered online, *not learners*. Learners who are not active participants in deciding the needs of new language classes may have little interest in in-class experiences of language learning and as a matter of course they fail to relate out-of-class experiences to their formal language environment. Such an observation enlightens a possible disconnect between the in-class experiences of language learners and their out-of-class experiences, particularly in such online spaces (Hafner, Chik, & Jones, 2015). Borrowing the term from Lotherington and Ronda (2014) communicative competence 2.0 demands, along with other competencies, the agentive participation which simply means learning by doing in class and out of class; accessing, joining, creating, sharing, remixing purposeful content. To accomplish these actions, learners need to gain information management strategies: finding online texts, assessing their worth, figuring out how to use them, and creating without stealing.

2. Defining Digital Literacy

The definition of literacy has gradually shifted from pen-paper to use of digital tools due to developing technology and ease of digital tool ownership. Martin (2005) suggests a comprehensive definition of digital literacies as "the awareness, attitude, and ability of individuals to appropriately **use** digital tools and facilities to **identify, access**, **manage, integrate**, **evaluate, analyze** and **synthesize** digital resources, **construct** new knowledge, **create** media expressions, and **communicate** with others, in the context of specific life situations, in order to enable constructive social action; and to **reflect** upon this process" (p. 135). Martin's definition reminds us of the procedure of critical thinking that consists of similar steps: identifying knowledge, comprehending, applying, analyzing, synthesizing, and evaluating information. In other words, being digitally literate mandates critical thinking for the purpose of making active and skillful use of new interactive platforms.

New Literacy Studies explain what digital literacy encompasses, because it is a broad term and point out the prominent aspects that educational contexts benefit from such as interactive environment, social networks and collaboration, learner-centeredness, creative reproduction of the digital materials etc. (Eshet-Alkalai, 2004; Hafner et al., 2015; Hoobs &

Tuzel, 2015; Mills, 2010, Tour, 2015). In language education, digital literacy receives attention as it is essential in and outside of institutional settings. Dooly and O'Dowd (2012) state that competent use of both FL skills and "e-skills" or "new literacies" helps learners be more able to work collaboratively in new online learning environments (p. 15). The new learning contexts offer a great amount of options with online affordances, and L2 learners have "various motives and goals in accessing L2 materials and in interacting with L2 communities" (Godwin-Jones, 2018, p. 8). For language teachers and learners, it is not new to utilize multimedia in language classes to include authentic and socio-culturally rich materials, but what is newer here is that online services often support online team collaboration to create, organize, and share information. L2 learners participate in multicultural and multilingual contexts where they are expected to collaborate with monolinguals, bilinguals, polyglots whose motive is to learn a foreign language. This interactive workspace is very rich in language and also carries risks of getting confused or lost for learners lacking the new digital literacies.

3. Digital Literacy Model and Popular and Prospective Tools for Teaching/Learning

Jones and Hafner (2012) provide a model for digital literacies that shows how the affordances of digital tools facilitate not only ways of meaning, but also ways of doing, relating, thinking, and being. *Doing* is using Instagram, Flicker, Pinterest etc. simply to share information, which shows how effectively we can manage pictures, texts, and videos in online spaces. *Meaning* is the form of representation, which teaches us to communicate effectively via web pages, social network sites, and related applications. *Relating* presents patterns of interaction, which users write on blogs, comment on Facebook, share thoughts in Twitter, contribute to Wiki with an online article written collaboratively to attract unknown audience etc. *Thinking* is more productive use of digital tools since it requires experiencing reality: participating oral or written communications. Ultimately, *being* comes to a level of acquiring a social identity that one presents oneself in a social network site to a certain community. This last dimension is more personal and sensitive and should be carefully considered when putting into practice in classroom spaces (Hafner et al., 2015, p. 2).

A new tool with different features arrives every day, thus, the border of digital literacy extends each day. This is a challenge for both teachers and learners to keep up with the new platforms. Currently, the popular online spaces are YouTube and Facebook. YouTube has been providing videos which teachers of English utilize-by adapting or simply playing- in language classes. However, going beyond being a consumer, now learners are asked to be producers in several ways. One can use iMovie or Windows Movie Maker to put some pictures, texts, and music together and upload it to this free online platform. One can also shoot a short video regarding, for example, attractions, events, news, instructions etc. and upload it to the created profile on YouTube. Moreover, making comments on available videos makes learners contributors as well. Facebook is relatively more used than any other free online platforms in terms of social aspects. Teachers have been asking how and more importantly why we use Facebook in their classes. How to use it varies depending on purpose and level of learners' digital literacy. However, the answers of why to use it may relate to the convenience of the medium as opposed to other media such as sharing documents or forming groups for class communication, but it might also be about how much teachers are willing to contemplate in the adaptation of social media to a traditional classroom setting. Since Facebook presents hyperlinks to many other platforms such as TedTalks, YouTube videos, Pinterest, Flicker, Wikipedia, BBC news, online shopping sites etc. which could be utilized to seek out potential sources for the use of language learning.

The emerging technology provides many platforms not only to consume information, but also evaluate, contribute, and create. Most of these platforms could be adapted to language teaching and learning. The major issue is 'how' to achieve bridging. For example, younger learners of language are into gaming and taking a significant part in social gaming communities. I have met a 20-year-old boy who speaks native-like English and told me he has never set foot in an English-speaking country in his entire life, but only played games and had written (discussion forums and blogs) and more frequently spoken conversations via League of Legends, Diablo and World of Warcraft. It is hard not to notice the source of such motivation and that gaming demands real communicative language. Moreover, the games are learners' own choices, personal engagement, and not forced by the school. Therefore, it would be practical to encourage gaming for learners to carry out further English studies out of class.

The new digital tools can take high-quality photos and videos instantly. This resulted in frequent use of applications like Snapchat or Instagram, which allow personalized sharing. Learners get proficient at the personal reflection, interpreting visuals, criticizing, remixing, and tagging. Nowadays, tagging is the most beneficial and practical way of finding *related* information among a vast amount of information available online now. Most tags are English-medium oriented. This could be taken as a great opportunity to prepare original activities for classrooms by using image-mapping to visualize and contextualize vocabulary teaching and help memory recall the words more easily.

Learners' motivation for digital inclusion shows how eager they are to explore and experiment, and how much further they are willing to go with digital literacy. They mostly learn to use tools for editing images, sound, and video. The editing tools became easy to access since they are now available in smartphones, tablets, and personal computers apart from desktops. They do not demand high-quality hardware or give a hard time using available storage as they used to do on desktops. For video editing VideoPad, iMovie or Windows Movie Maker, Adobe After Affects etc. are convenient software, which are user-friendly. For teachers, it is convenient to crop and edit the authentic materials for the level of learners or the length of the class/task. Furthermore, Screencast-o-matic allows to create a video of the actions of a computer screen. It can be used when teachers aim to give feedback on the writings of students on the interactive sheets as the video recording of the screen allows voice and camera recording and ultimately sharing of these videos.

Picture editing became dramatically easy via Instagram, Aviary, Typic etc., which are extremely easy to use unlike Photoshop, and cloud-based services that slow down the tool. Using realia in in-class tasks, assignments, portfolios, presentations etc. needs more than finding pictures through search engines. It entails appropriate placement of the pictures in the teaching/learning materials.

Voice Thread, Audacity, and GarageBand are only some examples of sound editing. Finally, some remixing tools such as TourBuilder, Popcorn Maker, and Aurasma can combine different sorts of sources (image, text, sound, and video) and produce rich materials (Godwin, 2015). HP5, a website to create interactive HTML5 content, allows to create, share, and reuse the materials. It can be used to create, for example, take-home speaking assignments where the teacher adds a picture and voice recording component with the instructions on an HP5 sheet and shares the link with the students. The voice recording component allows learners to play, download, listen to, and share their voices. The input, a picture etc., can enrich the materials/assignments and their output value. Moreover, PBWorks and Google Docs can also be used for educational purposes. Both provide an interactive

workspace for online team collaboration where teachers can invite the students to create a sheet collaboratively by using multimodal.

A controversial digital tool commonly used by language learners is the machine translation. Google, Microsoft, Yandex, Bing, and Babylon and similar web tools offer translations based on statistical equivalences between the language of text and target language. When learners use translation tools in writing a paper, an assignment etc., they should know that this kind of translation is less reliable and the percentage of accuracy regarding grammar and vocabulary is based on language pairs, the size of the dual language corpora, and even the use of punctuation. Therefore, teaching learners how to utilize machine translation is significant. They can use translation as a framework and make the necessary corrections of errors to have a faster translation. The machine translations are not very sensitive in terms of certain rules that are particular to one language. The errors in translations display the nature of a certain language as well as pointing to specific contrasts between learners' L1 and the target L2 (Godwin, 2015).

As Internet sources become massive, it might be the case that finding what we are looking for is getting harder due to the noise. When learners understand how to search on the Internet effectively, they discover that keywords, tags, hyperlinks, forums, discussions etc. allow them to locate the right online sources in seconds. There are different kinds of search engines to make the search more related. For example, visual search engines for words such as http://www.visualthesaurus.com/ and http://www.visuwords.com/ help language learners use dictionaries of a better kind. Also, image search engines such as http://taggalaxy.de/ could give learners a chance to look for visuals regarding a certain topic combined from several other search engines. A possible use can be that learners of a class can create their own visual dictionary based on the content of a group of lessons. Furthermore, it is quite easy to access free online dictionaries by Oxford, Cambridge, Longman, Macmillan, Langenscheidt and many other publishing companies or their applications for smartphones and tablets.

4. Online Platforms for Learners and Teachers

The availability and ease of access to digitals tools can make educators think that accessing is just the beginning. The important step is the readiness of individuals to use technology, communication networks, and information efficiently, effectively, and productively (Lemke, 2003). Douglas Rushkoff (1999) states in his book *Playing the Future,*

"Our kids are younger and less experienced than us, but they are also less in danger of becoming obsolete. They are the latest model of human being. Looking at the world of children is not looking backward at our own past-it's looking ahead. They are our evolutionary future (p. 4)."

Obviously digital natives, which is a term first coined by Prensky (2001) to describe learners born into technology, are fully engaged in everyday practices of technology. Many of them already know how to manage social media tools, online gaming, avatars, mixing videos, producing music, editing via digital tools, information hunting etc. These powerful sources can appeal to them in classroom in social, cognitive, and personal level as well. In that case, language classrooms ought to welcome multiple digital platforms to make environments authentic for learners such as Twitter as a platform to keep a micro-journal, Facebook as a class social networking tool, and YouTube as a video logging opportunity—virtually any digital platform can fit into the language classroom (Lotherington & Ronda, 2014). While all digital platforms are so ready to take place in classrooms, how is the situation with the language curriculum? Are curriculum, instruction, and assessment truly preparing students for the realities of today's communicative landscape (Guikema &

Williams, 2014)? On the one hand, most curriculums have to include the suggested Learning Management Systems such as Google Classrooms, Blackboard Learn, Moodle, or Edmodo which still help learners gain skills required for managing information; however, Godwin (2012) discusses that the spoon-fed content and closed environment LMS offers a more predictable and uniformed learning environment. On the other hand, most curriculums seem to deprive learners of collective working of multiple users around the universe, brainstorming, creating a solution for a mass problem etc. since it barely involves interactional dynamics among the learner groups in LMSs. To promote collective learning, curriculum, at least, could be flexible enough to permit teachers to establish a learning environment with digital content (by adapting or replacing the activities) which learners get to study on meaningful and real digital content. Also, the content may aim at writing for a global community (unknown audience) as in the case of Wikipedia, Flicker, blogs etc., joining game-related communities to improve both writing and reading, and employ a virtual identity which may keep learners away from any possible negative feelings that may occur in face-to-face situations or in the classroom.

A highly networked platform that supports a learner-centered pedagogy and non-traditional forms of teaching approaches is connectivist Massive Open Online Courses (cMOOCs) offered via popular platforms (Coursera, EdX, FUN, Futurelearn, MiriadaX etc.) (Yuan & Powell, 2013). It encourages inclusion since the ultimate success heavily depends on interaction via provided tools (hyperlinks, Google Hangouts on Air, Facebook clinics etc.). Digital literacy is important for learners' inclusion in these open courses when "consuming content, taking quizzes and exams, taking part in activities such as writing assignments and peer grading, and actively participate in discussions via discussion forums, blogs, twitter, Google+, or other forms of social media" (Hill, 2013). There is an increasing number of language MOOCs, and learners can benefit from these language courses provided by prestigious universities and institutions worldwide. Maintaining an active user profile in MOOCs demands developing digital competences and brings personalized learning environments. Various MOOCs can be made a part of the syllabus of such departmental courses as Information and Communication Technology (ICT) to achieve a blended learning.

Inclusion in such vast platforms as abovementioned brings about an issue of intercultural sensitivity. A useful tool to develop intercultural competence can be the Autobiography of Intercultural Encounters (AIE), which is a free source and part of the European Language Portfolio (ELP) (http://www.coe.int/t/dg4/autobiography/default_en.asp). As a diary designed to discover personal experiences, it can help reflecting both learners' intercultural encounters in person and in online platforms. For the encounters through visual media, the version available is An Autobiography of Intercultural Encounters through Visual Media (AIEVM) on the same website. As intercultural education is a sensitive area, which could easily emerge in global platforms with multiple users across the globe, it requires awareness of how societies and cultures differ and the fact that correctness, errors, attitudes, morality, values, beliefs, and so forth change from one to another.

5. Related Concerns

Different education cultures entail critical differences in motivation to use digital technologies. Teachers and learners may show resistance to digital in their own context. When the teachers are not provided with the knowledge and skills to integrate technology into their teaching through trainings, the motivation profiles will be affected negatively (Pouzevara, Dincer, Kipp & Sarnsik, 2014). Furthermore, the motivation might be age-related. The digital immigrants do not have as many experiences with digital tools as the digital natives. The teachers who are digital immigrants have their attitudes and beliefs

challenged when teaching in digitally-equipped (virtual) classes. However, lifelong learning centers, (virtual) workshops, and MOOCs about technology use could be an effective solution for those teachers to break the habit of their mind about digital tools.

Seeking information to write a paper is a meticulous activity, which schools usually fail or ignore to teach. It used to be encyclopedias and printed books that students based their assignments on while now the ease of finding information leads them to spend time on surfing between windows and be exposed to all sorts of information on websites. The first step is finding related information through right keywords, and then evaluating available information online to decide whether it is valid, related, and worthy. At this point of research, the main concern is judging the validity of the information presented on the website. Digital literacy requires learners to distinguish the validity of a source by evaluating where the source comes from: Is it a research paper, book, or journal with author and publisher information? Or is it simply Wikipedia, blogs, personalized papers etc. regarding the topic that may contain untrue or misleading information? In their research, Foster and Gibbons (2007) have found that students rely on a website by evaluating its popularity. This attitude will most likely lead to the use of incorrect and insignificant resources. Next step is note taking, which students usually copy and paste the information to make it coherent by either paraphrasing or using quotations. Paraphrasing an idea and integrating it to a paper within a cohesive text involve critical judgment. Since technology brings us varied sources, it is an opportunity to review the sources and generate new ideas out of them. Peters and Frankoff (2014) defines the process as constructive, creative, and productive since the text is constantly being amended by visiting and adding new information.

Socio-economic reasons interfere with the technology-enhanced learning/teaching as well. Easy access and availability of digital tools are not always an option for economically-challenged educational systems. On the other hand, political issues may restrict the access to once available open sources such as the Turkey case with Wikipedia where Turkey prohibited the use of Wikipedia to political concerns. Such cases often occur in Turkey with several social media tools as well as websites prohibited by Association of Access Providers in Turkey with public/private connections. In this case, it is important to be sensible about offline storage of the materials and information one has already worked on in order to avoid the loss of sources.

The significant issue at the point is *plagiarism.* Cheating on a work takes place in different forms: one can steal sentences, ideas, or a whole work to make it one's own. Owunwanne, Rustagi, and Dada (2010) have found out that plagiarism is acceptable and often carried out by students. The ease of accessing information may trigger the will to copy and paste instead of creating or producing when the mindset is not constructed. McGowan and Lightbody (2008) states that it is not surprising to discover that copyright issues are seldomly taught in high school or college, seeing how learners tend to plagiarize either intentionally or unintentionally. It is teachers' duty to educate learners about how to use sources, reference the author(s), and avoid cheating while paraphrasing or quoting. There are websites providing proper information of how to manage a certain style such as APA, MLA etc., which teachers can orient learners to. Moreover, Microsoft Word and Mendeley offers the citation use and automatic bibliography for the convenience of academics, learners, or simply anyone writing papers in a formal manner.

An important issue to deal with when teaching digital literacy is filter bubble. The algorithms used by Google and other search engines tend to navigate users to the sources related to their previous searches, products purchased, geographical location, and personal information gleaned, such as age, profession, or marital status (Godwin, 2015). This simply

means that no two people's results of a search will be the same, even with the same keywords or tags. One the one hand, it may sound practical since it seems to eliminate information we may not be interested at all. On the other hand, these filters step in our searches and decide what to see and not to see on behalf of us. As Kern (2014) states the concern that the World Wide Web may be becoming less and less of a window onto the world and more and more of a window onto ourselves. Each single click by us determines our future clicks and sentences us to more restricted and personalized windows.

A possible factor that may cause teachers to reconsider including digital contents is that learners may encounter various kinds of language in informal online spaces, and this may lead the heavy use of code switching, hybridized forms, slang, specialized vocabulary, and formulaic speech (Godwin, 2015). On the other hand, there would be no drastic harm in introducing informal language, different registers and styles, and idiomatic expressions unless the content consists of unethical, racist, and interculturally insensitive language use.

5. Conclusion

The promise of digital tools to enhance language teaching and learning has been deemed. Therefore, the matter of effective and skillful use of various digital tools has gained significance. New literacy studies emerged after the arrival of interactive tools and platforms provided free-of-charge. The common idea in these studies is to improve digital literacy besides new skills. The use of multimedia, which requires more strategic skills such as interaction, collaborative communication, and agentive participation, has become more important than using passive materials (materials that are to consume). Thus, being digitally literate extended to a broader definition where it entails identifying, accessing, analyzing, managing, integrating, evaluating, and synthesizing digital tools for a particular purpose.

Learner-oriented teaching has also been promoted with digital literacy. The notion of mere consumption of the given materials has been abandoned, and a notion that promotes creativity with the inclusion of learners for accomplishment of learners has been adopted. This led to redefining the roles of teachers and learners.

The availability of tools like desktops, personal computers, tablets etc. today provides a mobile learning experience. It has become convenient to create knowledge and circulate it. In language education, the richness of multimedia can contribute to the learning through the use of popular social media tools, online gaming chat rooms, tagging tools, picture, voice, and video editing tools, mixing tools, interactive HTML5 contents, machine translation tools, search engines and (visual) thesauruses. This is a straightforward step in improving the effectiveness of learning materials such as tasks, assignments, portfolios, projects, feedback giving, presentations etc. Furthermore, platforms such as LMSs and MOOCs that function on multimedia necessitate good level of digital literacy. Similarly, as the platforms employ a learner-centered pedagogy, gaining digital competence is even more essential.

Lack of digital literacy brings controversial issues in educational contexts. The common concerns including resistance to use of digital tools, evaluating the validity of a source, bans of websites due to country policies, filtering problem, and insensitive language in interactive platforms. It hinders autonomous use of technology for educational purposes. In this case, lifelong learning centers, MOOCs, and (online) workshops can encourage teachers and learners to grasp a better understanding of digital competence to enhance their own language teaching and learning.

References

Dooly, M., & O'Dowd, R. (Eds.). (2012). *Researching online foreign language interaction and exchange: Theories, methods and challenges.* Bern, Switzerland: Peter Lang.

Eshet-Alkalai, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, *13* (1), 93.

Foster, N. F. & Gibbons, S. (2007). *Studying students: The undergraduate research project at the University of Rochester.* Chicago, IL: Association of College and Research Libraries.

Godwin-Jones, R. (2012). Challenging hegemonies in online learning. *Language Learning & Technology, 16* (2), 4–13.

Godwin-Jones, R. (2015). Contributing, creating, curating: Digital literacies for language learners. *Language Learning & Technology, 19* (3), 8–20.

Godwin-Jones, R. (2018). Chasing the butterfly effect: Informal language learning online as a complex system. *Language Learning & Technology*, *22* (2), 8–27.

Guikema, J. P., & Williams, L. (2014). Digital literacies from multiple perspectives. In J. P. Guikema, & L. Williams, *Digital literacies in foreign and second language education* (pp. 1-7). The USA, Texas: CALICO.

Hafner, C. A., Chik, A., & Jones, R. H. (2015). *Digital literacies and language learning. Language Learning & Technology*, *19* (3), 1–7.

Hill, P. (2013, March 10). Emerging student patterns in MOOCs: A graphical view. *E-Literate*. Retrieved from http://mfeldstein.com/emerging_student_patterns_in_moocs_graphical_view

Hoobs, R. & Tuzel, S. (2015) Teacher motivations for digital and media literacy: An examination of Turkish educators. *British Journal of Educational Technology*, *48* (1), 7-22.

Jones, R. H., & Hafner, C. A. (2012). *Understanding digital literacies: A practical introduction.* London, UK: Routledge.

Kern, R. (2014). Technology as pharmakon: The promise and perils of the internet for foreign language education. *Modern Language Journal, 98* (1), 340–357.

Lemke, C. (2003). The North Central Regional Educational Laboratory and the Metiri Group. Retrieved on January 1, 2016, from www.ncrel.org/engauge: http://pict.sdsu.edu/engauge21st.pdf

Lotherington, H., & Ronda, N. (2014). 2B or Not 2B? From pencil to multimodal programming: New frontiers in communicative competencies. *Digital Literacies in Foreign and Second Language Education* (p. 9). CALICO.

Martin, A. (2005). DigEuLit—a European framework for digital literacy: A progress report. *Journal of eLiteracy*, *2,* 130-136.

McGowan, S., & Lightbody, M. (2008). Enhancing students' understanding of plagiarism within a discipline context. *Accounting Education*, *17*, 273–290.

Mills, K. A. (2010). A review of the "Digital Turn" in the new literacy studies. *Review of Educational Research*, *80* (2), 246-271.

Owunwanne, D., Rustagi, N., & Dada, R. (2010). Students' perceptions of cheating and plagiarism in higher institutions. *Journal of College Teaching and Learning*, *7*, 59–68.

Peters, M., & Frankoff, M. (2014). New literacy practices and plagiarism: A Study of Strategies for Digital Scrapbooking. In J. P. Guikema, & L. Williams (Eds.), *Digital Literacies in Foreign and Second Language Education* (Vol. 12, pp. 245-264). The USA, Texas: CALICO.

Pouzevara, S., Dincer, A., Kipp, S. & Sarnsik, Y. (2014). *Turkey's FATIH project: A plan to conquer the digital divide or a technological leap of faith?* Sabanci University, Education Reform Initiative (ERI) and Research Triangle Institute (RTI), Turkey.

Prensky, M. (2001). Digital natives, digital immigrants. Part 1. *On the Horizon*, *9* (5), 1–6.

Rushkoff, D. (1999). *Playing the future: What we can learn from digital kids aka children of chaos.* New York: Riverhead Books.

Stockwell, G. (2015). Digital media literacy in language teaching. In *The 17th International Conference, 서울대 국어교육연구소*.

Tour, E. (2015). Digital mindsets: Teachers' technology use in personal life and teaching. *Language Learning & Technology, 19* (3), 124-139.

Tour, E. (2012). TESOL in times of change. *Monash University Linguistics Papers*. Special Issue, *8* (1), 11–21.

Yuan, L. & Powell, S. (2013). *MOOCs and open education: Implications for higher education.* Centre for educational technology & interoperability standards (Cetis). Retrieved from http://publications.cetis.org.uk/wp-content/uploads/2013/03/MOOCs-and-Open-Education.pdf

THE ROLE OF SUBTITLES IN FOREIGN LANGUAGE TEACHING

Bengü Aksu Ataç

Nevşehir Hacı Bektaş Veli University

aksubengu@gmail.com

Sevtap Günay-Köprülü

Nevşehir Hacı Bektaş Veli University

sgkoprulu@hotmail.com

Bengu Aksu Ataç is the director of the school of Foreign Languages and Head of the ELT programme. Her professional interest areas are language teaching, testing and assessment, Common European Framework of Reference for Languages, European Language Portfolio, authentic assessment, peace education and ELT for young learners.

Sevtap Günay Köprülü is an Assistant Professor in Department of Foreign Language Education in Nevşehir Hacı Bektaş Veli University.

THE ROLE OF SUBTITLES IN FOREIGN LANGUAGE TEACHING

Bengü Aksu-Ataç

aksubengu@gmail.com

Sevtap Günay-Köprülü

sgkoprulu@hotmail.com

Abstract

Today, engaging in real language environments and exposing to the target culture often are seen as more appropriate to learn a second language. In this sense, movies in foreign language teaching are important audio-visual tools in terms of reflecting the language, culture, customs and traditions, clothes and lifestyle of a community. In language teaching, it is very important to know how words and expressions are used in different contexts and to pronounce them correctly. Many language learners benefit from these audio-visual tools because of the contributions of movies to foreign language learning. Especially in the development of listening comprehension skills, learners prefer to watch movies in foreign language. There are three options: watching the movie in its original language without subtitles, watching the movie in its original language with subtitles in source language and watching the movie in its original language through subtitle translation in target language.

In this study, 68 English Language Teaching prospective teachers have been investigated for their attitudes towards developing their English language skills outside of the educational setting, whether they use English movies to improve their language skills, which genre they choose and their subtitle preferences. Therefore, the factors that should be considered in the selection of movies to be used in foreign language learning, the subtitle as a method of movie translation, the place and importance of subtitles in language learning, the positive and negative aspects of movie translation with the subtitle method have been studied and some suggestions have been made. As a research tool, appropriate sampling questionnaire has been distributed to the students and the collected data has been analyzed by SPSS program. As a result of the research, students believe that watching movies has improved their listening comprehension skills and has a positive contribution to the development of their speaking skills. Thus, they prefer English subtitles more.

Keywords: subtitle, subtitle translation, foreign language learning

1. Introduction

Learning a language means communicating by acquiring the communicative competency of that language. This requires learning the communicative skills of the language, as well as productive skills such as speaking and writing and perceptive skills such as listening and reading, and also learning to communicate verbally and in writing with the native speakers of target language along with the culture, lifestyle, and behaviour of the country where the language is spoken (Edmonson & House, 1993, p. 82, as cited in Aktaş, 2005, p. 93). The most effective way to achieve this is to learn the language in the target country. However,

since such an opportunity is not always possible, benefiting from the movies in this sense is an appropriate approach. Movies are the most effective audiovisual tools to learn the culture of the target language, because you can both see and hear, moreover, witness the patterns and contexts of expressions and words. Language becomes complete only when it is learned together with the culture in which it is born.

There are many studies regarding the significance of movies in foreign language learning. It has been proved that it has a booster feature in teaching level in a short time. (Demirezen, 1990, p. 294). Although they are fictional, movies reflect real life. Learning a foreign language in a system that is close to real environments attracts attention of the learners and they can see the shapes of the theoretical knowledge about the language in the movies. Movies are important tools especially in the acquisition of listening comprehension and speaking competences because through movies one can learn about how to form dialogues in different environments, gestures and mimics supporting verbal expressions, formal and informal expressions, stress and intonation. Since movies are both aural and visual tools, watching a movie in a foreign language to learn a language facilitates both channels. For this reason, learning with two channels increases learning speed and its permanence. Chan and Herrero (2010, p.11) explain the advantages of using movies in foreign language teaching as follows;

1 The movies make understanding activities easier because they are perceived as real.

2 Movies create an atmosphere of curiosity with the introduction of mutual ideas and interpretations.

3 Movies help to explore nonverbal linguistic elements.

4 They develop verbal and aural skills.

5 Movies help learners to speak more naturally and fluently, create more meaningful texts and improve their word power.

6 There are many ways to use movies in language classes depending on their genres.

It is apparent that they are very useful when considering all these advantages of movies in foreign language learning and in terms of developing listening-comprehension and speaking skills it is also obvious that movies should be watched in their original language. In such situations using subtitles will further motivate learners in the first stages of language learning. When the movie is watched in its original language, as the language in movies reflects their real use in the natural environment, learners can be aware of the use of linguistic features such as emphasis and intonation. It becomes easier and faster to learn the language as it integrates what is seen with what is heard. On the other hand, watching a movie in a foreign language is more difficult and tiresome because it can be difficult to understand. This may negatively motivate language learners. However, subtitles can be used to prevent such an inconvenience.

1.1. Place of Subtitled Movies in Foreign Language Learning

In terms of observing vocabulary learning, listening comprehension, differences in expressions, and especially pronunciation, subtitled movies are suggested to language learners. Studies held in a research institution in England show that presenting audio with the written text at the same time has positive effects on recording spoken words and increasing language memory (Okyayuz, 2016, p. 43). In terms of foreign language learning, the most important advantage of subtitle is that the movie can be watched in its original language and with the original audio. Therefore, audience/language learner, who has partial knowledge of the language, can benefit from the subtitle when s/he cannot understand the audio.

Subtitles used in movies can be classified under various types. For example, bimodal, standard and reversed subtitling (Zanon, 2006, p. 47). However, there are two different types of subtitles in general:

1.Interlingual subtitle: The original language of the movie is translated into another language. The audience/language learner watches and listens to the movie with the original/target language, but reads the subtitle translated into his/her own language.

2. Intralingual subtitle: Here, the original language and the subtitle of the movie are the same. Intralingual subtitle is more for people with impaired hearing. Therefore, with this kind of subtitle, all audio sounds like footsteps, which are a sign of someone is coming, knocking sound at the door, the sound of a car horn that is coming from outside are shown. There are also intralingual subtitles for foreign language learners apart from the aforementioned ones.

As a result of many research, it is seen that using subtitles in foreign language learning has positive results. These positive results also include subtitles' positive effects on productive skills like recalling the words entirely and keeping them in the correct context besides communicative performance in a specific oral and written communication (Rokni & Ataee, 2014, p. 718).

In a research conducted, it has been decided that subtitles in the original language (English) have displayed positive effects on listening comprehension skills of students, who are learning English as a foreign language, with medium level English competency. The use of subtitled movie has been found helpful and entertaining by the participants (Felek Başaran, 2011). In another study, it has been identified that listening comprehension skills of the audience/language learners that watch a movie in English with intralingual subtitle have improved more than the others that watch it without any subtitle (Rokni & Ataee, 2014).

In 2009, in an extensive research carried out on the use of subtitle in language education, it is revealed that intralingual and interlingual subtitles have positive contributions to the foreign language education. This paper has demonstrated that subtitles increase the motivation in language education and support a multilingual environment, and learners that want to improve their language skills prefer subtitle translations instead of dubbing translations (EACEA, 2009/01, as cited in Okyayuz, 2016, p. 43).

Many researchers agreed that intralingual subtitles have positive effects in foreign language learning. Studies held among English language learners have revealed that intralingual subtitles can be effective as a teaching tool (see Okyayuz, 2016, p. 43). In many research, it is stated that subtitles have positive influences on language education in general terms, however, inter-lingual subtitle, which is translated subtitle, has been recommended less than intra-lingual subtitle, which is the subtitle in the original language, by comparison. As an example, according to the study carried out by Holger Mitterer and James McQueen (2009), a learner who watches a movie in a foreign language to improve listening comprehension skills in that foreign language can ideally achieve this by watching the movie with an intra-lingual subtitle. Similarly, Caimi (2006) states that intra-lingual subtitles can be used for didactic purposes in language learning. In another paper, a movie in English with English subtitles was watched and it was seen that the subtitle helped students to learn new words and idioms, and to improve their listening skills (Özgen, 2008).

The main goal of interlingual subtitle is to use expressions, which make the movie understandable. The audience/learner comprehends the meaning of the scene by reading the subtitle. On the other hand, there are also features that subtitles should have technically (Günay-Köprülü, 2016). Because of these technical features like the number of lines and the

number of maximum characters, translation of the subtitles has to be restricted, therefore subtitles, in a way, have to give the summary or the gist of spoken expressions.

2. Methodology

In this study, the first and second year students of English Language Teaching department at Haci Bektas Veli University in Nevsehir have been investigated for their attitudes towards developing their English language skills outside of the educational environment, whether they use English movies to improve their language skills, which genre they choose and their subtitle preferences. As a research tool, appropriate sampling questionnaire has been distributed to the students and the collected data has been analyzed by Statistical Package for Social Sciences (SPSS) program version 18.0 to calculate means, frequencies and percentages.

The questionnaire has been constructed in order to have general information about participants, their opinions and attitudes towards watching movies, their preference of subtitles and the effects of movies and subtitles on language learning skills. The opinions of the participants on the effects of movies and subtitles on language learning skills have been graduated by Likert Scale: Strongly disagree = 1, Disagree = 2, Neither agree nor disagree = 3, Agree = 4, Strongly agree = 5.

While identifying the study and the questionnaire, "A survey of students' opinions on watching English soundtrack movies to enrich listening skill development" by Tanisara Pimsamarn has been used and adapted.

3. Findings and Discussion

The first and second year students of English Language Teaching department at Haci Bektas Veli University in Nevsehir have been participated in the study. 34 first year students, 34 second year students and total 68 students have participated in the study. 86.8 % of the 68 students were (N=59) girls and 13.2% (N=9) of them were boys.

Only 2 of the 68 students participated in the questionnaire did not answer the question asking about the type of graduate high school. Over 66 students who have answered the question 78.8 % (N=52) of them were Anatolian High School graduates, and 21.2% (N=14) of them were Normal State High School graduates.

Only 2 of the 68 students participated in the questionnaire did not answer the question asking about their duration of language learning. Over 66 students who have answered the question 74.2 % (N=49) 11-15 years, 21.2% (N=14) 6-10 years, 3% (N=2) 1-5 years and only 1 student has more than 16 years of English language education period.

29.4% (N=20) of the students have mentioned that they watch a movie once a week and again 29.4 % (N=20) them said they watch a movie once a month. While 29.4% (N=20) of the students have been watching an original English movie, 27.9% (N=19) pointed that they watch original English movies less than a month. For more detailed information see Table 1.

Table 1. *Movie watching frequency (MWF) and English movie watching frequency (EMWF)*

	Opportunity to use Movies/ English Movies	Frequency		Percent		Valid Percent		Cumulative Percent	
		FİS	İFİS	FİS	İFİS	FİS	İFİS	FİS	İFİS
Valid	Everyday	3	2	4.4	2.9	4.4	2.9	4.4	2.9
	Three to five times a week	15	10	22.1	14.7	22.1	14.7	26.5	17.6
	Once a week	20	17	**29.4**	25.0	29.4	25.0	55.9	42.6
	Once a month	20	20	**29.4**	**29.4**	29.4	29.4	85.3	72.1

Less than once a month	10	19	14.7	**27.9**	14.7	27.9	100.0	100.0
Total	68	68	100.0	100.0	100.0	100.0		

48 students answered the question about what kind of movies they watch. The percentage of their answers: 27.1 % (N=13) Comedy, 22.9 % (N=11) Action, 18.8% (N=9) Romantic, 10.4% (N=5) Thriller and the rest have preferred Other.

56 students answered the question about why they watch movies. 83.9 % (N=47) of them mentioned that they watch movies in order to improve language skills and 8.9 % (N=5) for fun.

A great number of students (%94.0) who have participated in the questionnaire prefer subtitles while watching movies (see Table 2). About the languages used in subtitles although there is not a big difference between languages the number of the students who prefer English subtitles (% 53.2) are more than the others (see Table 3).

Table 2. *Subtitle preference*

	Subtitle	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Yes	63	92.6	**94.0**	94.0
	No	4	5.9	6.0	100.0
	Total	67	98.5	100.0	
Missing	99.00	1	1.5		
Total		68	100.0		

Table 3. *Language of the subtitle*

Subtitle language		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Turkish	29	42.6	46.8	46.8
	English	33	48.5	**53.2**	100.0
	Total	62	91.2	100.0	
Missing	99.00	6	8.8		
Total		68	100.0		

50 of the 68 students have answered the question about why they prefer English subtitles while watching movies. 46% (N=23) of 50 students prefer English subtitles in order to understand the dialogues while watching movies. Again 46% (N=23) of the students prefer English subtitles to see the English equivalents of the words they hear (see Table 4).

Table 4. *Reason for English subtitle preference*

	English subtitle	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Infer the meaning from the written form	23	33.8	**46.0**	46.0
	To see the written forms of the words	23	33.8	**46.0**	92.0
	Others	4	5.9	8.0	100.0
	Total	50	73.5	100.0	
Missing	99.00	18	26.5		
Total		68	100.0		

53 students have answered the question about why they prefer Turkish subtitle while watching a movie. 50.9 % (N=27) of the 53students mentioned that they prefer Turkish subtitle to see the Turkish equivalent of the words in the movie, 24.5% (N=13) of them to understand the movie, and 22.6% (N=12) of them to understand the dialogues in the movie (see Table 5).

Table 5. *Reason for Turkish subtitle preference*

	Turkish subtitle	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Understand the movie	13	19.1	**24.5**	24.5
	Understand the dialogues	12	17.6	**22.6**	47.2
	To see the Turkish equivalance of the word	27	39.7	**50.9**	98.1
	Other	1	1.5	1.9	100.0
	Total	53	77.9	100.0	
Missing	99.00	15	22.1		
Total		68	100.0		

According to the results of the questionnaire, most of the students agree with 10 of the 11 questions. The only question that they do not agree and confused is the one about the Turkish subtitles of the foreign movies. The students are indecisive about the adequacy of the Turkish subtitles. They totally agree with the following opinions; Learn correct pronunciation through movies, By watching movies in original English language I learn new words, I can improve my listening skill by watching English language movies (see Table 6).

Table 6. *Opinions on the effects of movies and subtitles on language learning skills.*

	N	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1. English subtitles help me understand conversation in the movies better	68	-	1 (% 1.5)	5 (% 7.4)	**43 (% 63.2)**	19 (% 27.9)
2. Learn correct pronunciation through movies	68	-	-	5 (% 7.4)	31 (% 45.6)	**32 (% 47.1)**
3. Movies help me understand unfamiliar accents and dialects better	68	-	-	7 (% 10.3)	**36 (% 52.9)**	25 (% 36.8)
4. By watching movies in the original English language I learn new words	68	-	1 (% 1.5)	3 (% 4.4)	30 (% 44.1)	**34 (% 50)**
5.By watching English movies with Turkish subtitles I learn new words	68	-	6 (% 8.8)	7 (% 10.3)	**29 (% 42.6)**	26 (% 38.2)
6. I find Turkish subtitle translations sufficient	66	3 (% 4.5)	15 (% 22.7)	**34 (% 51.5)**	9 (% 13.6)	5 (% 7.6)
7. Movies help me understand unfamiliar idioms and slangs better	68	-	3 (% 4.4)	10 (% 14.7)	**34 (% 50)**	21 (% 30.9)
8. Original English movies can improve me listening skill better than other English media normally used in class	67	-	-	10 (% 14.9)	**35 (% 52.2)**	22 (% 32.8)
9.Original English movies is good for teaching listening for students	67	-	-	4 (% 6)	**32 (% 47.8)**	31 (% 46.3)
10. Watching English movies can be beneficial for English use in everyday life	68	-	-	2 (% 2.9)	**37 (% 54.4)**	29 (% 42.6)
11. I can improve my listening skill by watching English language movies	68	-	1 (% 1.5)	-	30 (% 44.1)	**37 (% 54.4)**

The idea which the students mostly agree with is that they can improve their English listening skills by watching original English movies. Secondly, they think that they can learn new words while watching original English movies. Thirdly, they agree that original English movies are so helpful in teaching listening skills to the students. The idea that they least agree with is that they can learn new words while watching Turkish subtitled original English movies. In addition, they find Turkish subtitles insufficient (see Table 7).

Table 7. *The average of the opinions on the effects of movies and subtitles on language learning skills.*

	N	Mean	Min	Max	Std	Variance
1. English subtitles help me understand conversation in the movies better	68	4.17	2.00	5.00	.621	.386
2. Learn correct pronunciation through movies	68	4.39	3.00	5.00	.626	.392
3. Movies help me understand unfamiliar accents and dialects better	68	4.26	3.00	5.00	.637	.406
4. By watching movies in the original English language I learn new words	68	**4.42**	2.00	5.00	.653	.427
5.By watching English movies with Turkish subtitles I learn new words	68	4.10	2.00	5.00	.916	.840
6. I find Turkish subtitle translations sufficient	66	2.96	1.00	5.00	.927	.861
7. Movies help me understand unfamiliar idioms and slangs better	68	**4.07**	2.00	5.00	.797	.636
8. Original English movies can improve my listening skill better than other English media normally used in class	67	4.17	3.00	5.00	.672	.452
9.Original English movies is good for teaching listening for students	67	**4.40**	3.00	5.00	.604	.365
10. Watching English movies can be beneficial for English use in everyday life	68	4.39	3.00	5.00	.550	.303
11. I can improve my listening skill by watching English language movies	68	**4.51**	2.00	5.00	.585	.343

4. Conclusion

Researchers agree that using movies has positive effects on foreign language teaching and learning. It is also emphasized that watching movies in the target language with subtitles motivates learners. Following these ideas, the students' attitudes and views on watching subtitle movies and whether they use movies for foreign language learning outside the classroom environment have been assessed. The evaluation was done according to the data obtained through the questionnaire which is one of quantitative research designs. The majority of the 68 students who participated in the survey graduated from Anatolian high schools. They have English education for 11-15 years. These students usually watch various movies once a week or once a month. The frequency of watching an original English movie is once a month. The vast majority of participants prefer to watch comedy and action movies. The majority watch movies for language development and they prefer subtitles. Although there is no big difference between the English-Turkish subtitles preference, English subtitles are preferred because they want to understand the dialogues and see spellings. Turkish subtitles are preferred in order to understand the Turkish meanings of the words they hear.

Students agree that movies help them learn accents and dialects, idioms, proverbs, and slang words in the target language and they also learn correct pronunciation. Opinions that students most agree with are that they can learn new words by movies and they can improve their listening skills. They think English movies are very helpful both in everyday English

use and in teaching English to students, especially to teach listening comprehension skills. The rate is a bit lower, considering that English movies are more helpful in improving listening and comprehension skills than other classroom tools. Regarding the use of subtitles in movies, compared to the view that they learn new words with Turkish subtitles, they agree more on the item that English subtitles are better for understanding dialogues in movies. According to results of the questionnaire, students are not sure about sufficiency of Turkish subtitles. This explains why the rate is lower for the item that Turkish subtitles are useful in new vocabulary learning.

As a result, students have the opinion that it is useful to watch English movies to improve their English language skills and it is useful to watch English movies with English subtitles.

References

Aktaş, T. (2005). Yabancı dil öğretiminde iletişimsel yeti [Communicative competence in foreign language teaching]. *Journal of Language Linguistic Studies*, *1*(1), 89-100.

Bird, S. A. & J. N. Williams (2002): The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics*, *23*, 509-533.

Caimi, A. (2006). Audiovisual translation and language learning: The promotion of interlingual subtitles. *The Journal of Specialised Translation, 6*, 85-98. Retrieved from www.jostrans.org/issue06/art_caimi.php

Chan, D., & Herrero, C. (2010). *Using movies to teach languages: A teachers' toolkit.* (It was produced for Cornerhouse (part of Projector: Community Languages). Retrieved on January 12, 2015 from http://www.cornerhouse.org/movie/movie-media/toolkit-for-teachers-using-movie-to-teach-languages

Demirezen, M. (1990). Video kullanımının yabancı dil öğretimine getirdikleri [The conributions of video use to the foreign language teaching]. *Hacettepe University Journal of Education*, 5, 289-298.

EACEA (2009): *Study on the use of subtitling: The potential of subtitling to encourage foreign language learning and improve the mastery of foreign languages*. European Commission, Directorate – General Education and Culture, Final Report June 2011, Media Consulting Group.

Felek- Başaran, H. (2011). *The effects of captioning on EFL Learners' listening comprehension: An investigative study with intermediate level university prep-school students*. (Master's thesis). Anadolu University, Eskişehir, Turkey.

Günay-Köprülü, S. (2016). Altyazı Çevirmenini Kısıtlayan Etkenler [The factors limiting the subtitle translator]. In H. Göksu (Ed.). *Eğitim bilimlerinde güncel yaklaşımlar* [*Current approaches in educational sciences*]. (pp. 1-10). Saarbrücken: LAP Lambert Academic Publishing.

Okyayuz, A.Ş. (2016). *Altyazı Çevirisi [Subtitle translation]*. Ankara: Siyasal Kitabevi

Özgen, M. (2008). *The use of authentic captioned video as listening comprehension material in English Language Teaching* (Master's thesis). Selçuk University, Konya, Turkey.

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *Plos One Max-Planck-Institut für Psycholinguistik*. Retrived from http://dx.doi.org/10.1371/journal.pone.0007785

Pimsamarn, Tanisara (2011). *A survey of students' opinions on watching English soundtrack movies to enrich listening skill development* (Master's thesis). Thammasat University Bangkok, Thailand.

Rokni, S.J.A., & Ataee, A. J. (2014). Movies in EFL classrooms: With or without subtitles. *The Dawn Journal, 3*(1), 715-726

Vanderplank, R. (1988): The value of teletext subtitles in language learning, *ELT Journal, 42* (4), 272-281.

Zanon, N. T. (2006). Using subtitles to enhance foreign language learning. *Porta Linguarum 6*, 41-52

NEW TECHNOLOGIES IN MICROTEACHINGS: IS IT POSSIBLE FOR WEB 2.0 TOOLS TO AFFECT PROSPECTIVE TEACHERS' WEB PEDAGOGICAL CONTENT KNOWLEDGE AND SELF-EFFICACY?

Halit Karalar

Muğla Sıtkı Koçman University

halit@mu.edu.tr

Bilge Aslan Altan

Muğla Sıtkı Koçman University

bilgeaslan@mu.edu.tr

Halit Karalar is an Assistant Professor in Computer Education and Instructional Technologies Department in Muğla Sıtkı Koçman University. He is interested in differentiated technology integration in all levels of education, e-learning, instructional design, and programming instruction.

Bilge Aslan Altan is a Research Assistant in Curriculum and Instruction Department in Muğla Sıtkı Koçman University. She is currently working on her dissertation. She is interested in curriculum adaptation, instructional designs, and instructional technologies.

NEW TECHNOLOGIES IN MICRO TEACHINGS: IS IT POSSIBLE FOR WEB 2.0 TOOLS TO AFFECT PROSPECTIVE TEACHERS' WEB PEDAGOGICAL CONTENT KNOWLEDGE AND TEACHER SELF-EFFICACY?

Halit Karalar

halit@mu.edu.tr

Bilge Aslan Altan

bilgeaslan@mu.edu.tr

Abstract

The purpose of this study was to determine the effect of Web 2.0 tools applied in microteachings on prospective teachers' web pedagogical content knowledge and teacher self-efficacy, and to examine their views on the procedure. This research used mixed research design and was conducted with 24 prospective teachers who participated in teacher training certificate program in Faculty of Education in a state university in Turkey. Web pedagogical content knowledge scale, teacher self-efficacy scale, and an open-ended question form were used as data collection tools. Paired samples t-test and Pearson correlation analysis were applied for the analysis of quantitative data; content analysis was used to analyze the qualitative data. The findings showed that there was a significant difference on prospective teachers' web pedagogical content knowledge and teacher self-efficacy levels after Web 2.0 tools implementation. The research also concluded that there was a medium level positive relation between teacher candidates' web pedagogical content knowledge and teacher self-efficacy. They stated that it was easy to use these tools, learning was fun, easy, and permanent, and they actively involved in instructional time. They added that they were eager to apply Web 2.0 tools in their future classrooms.

Keywords: microteaching, web 2.0 tools, web pedagogical content knowledge, teacher self-efficacy

1. Introduction

In our present world, policy makers and educationalists see the technology as a key tool for solving problems in an education system and so that the classrooms are being equipped with new technologies (Akcaoglu, Gumus, Bellibas, & Boyer, 2015). The pressure on teachers expected to use these new technologies has been increasing, and prospective teachers' training to use these technologies in future has come to topic (Tondeur et al., 2012; Tsai & Chai, 2012). In this study, microteaching method with Web 2.0 tools, which is an effective way to contribute to prospective teachers' training, was focused.

1.1. Microteachings

It has been remarkably highlighted that prospective teachers should experience microteachings so as to apply technology in their future classrooms effectively (Cavin, 2008; Ertmer & Ottenbreit-Leftwich, 2010). With microteachings, it is aimed to bring prospective teachers critical teacher behaviors that are determined beforehand. In this method, the

number of the students and the instructional time are decreased, and teaching activities are increased (Görgen, 2003). According to Çoban (2015), generally, microteaching method is a teaching experience which is comprehensive and in which it is revised and reevaluated. Microteachings are the first steps prospective teachers to make before they face with real classroom teachings. These teachings are of key procedures to guide prospective teachers to determine their course objectives, to design their instructional methods, to select their materials, to maintain a course hour, and to provide feedback for their own teachings.

1.2. Self-Efficacy

Self-efficacy issue has been brought forth in technology integration in education in many studies (Curts, Tanguma, & Peña, 2008; Niederhauser & Perkmen, 2008). As these researchers highlight, teachers' context and technology knowledge increase simultaneously with technology-integrated instructions, therefore, teachers gain more confidence to adapt and use technology professionally. Meanwhile, self-efficacy is explained by Bandura (1977) mostly as individuals' beliefs shaped by their own motivation, affection, interaction, models and many other dynamics. Self-efficacy for a teacher is a way of fulfilling a duty in a specific environment and obtaining an outcome as a conclusion (Bandura, 1986). To sum up, teachers' self-efficacy affects their way of instruction through their skills, habits and competencies. Regarding self-efficacy, teachers can feel more confident to maintain a process on internet and digital world. Besides cherished with high self-efficacy, teachers' motivation is likely to increase in interacting with internet and its agents (Horzum, 2011).

1.3. Web 2.0 Tools

Tim O'Reilly first introduced the concept of Web 2.0 in 2004 (O'Reilly, 2007). Though there has not been a consensus on the clear definition of Web 2.0, it is widely accepted that there is a transition from Web 1.0, which is a solely read-only web, to a new era in which web can be read-written thanks to Web 2.0 and that the users' role has been modified (Hew & Cheung, 2013). In Web 2.0 era, individuals who were passive users in Web 1.0 era have become active participants of information and individual or collaborative producers of content (Brown, 2012; Hew & Cheung, 2013; Solomon & Schrum, 2007). The change in users' roles has resulted from Web 2.0 tools.

Web 2.0 tools are of the easiest agents of digital world to apply in education (Kurilovas & Juskeviciene, 2015). Students of our age benefit from these tools' free of charge, easiness to use (Churchill, 2011), being user-friendly, promoter of learning anywhere-anytime, and so these tools encourage individual and/or collaborative learning and production (Brown, 2012; Hartshorne & Ajjan, 2009). These technologies remarkably take turns on education. As students utilize, teachers also take advantage of Web 2.0 tools. They can share their experiences and ideas on a specific topic with their colleagues all around the world. It is feasible to follow a further educational career, webinars, workshops, or any applicable lectures. On the other hand, these tools can provide teachers with instructional strategies, activities, materials, assessment alternatives, shared platforms for students, or any applications that can be handy in classrooms. In addition to these tangible contributions, teachers can foster inner qualities as well.

Studies (Laru, Näykki, & Järvelä, 2012; Meyer, 2010) have presented different results that Web 2.0 tools have contributed to learners and individuals' different needs. These results showed that Web 2.0 tools have increased collaboration, effective decision-making and individual knowledge acquisition through interactive participation. Besides, individuals' Web 2.0 tools knowledge and their intention to apply were found to be affected highly by their preferred learning styles, beliefs about worthiness of these technologies for improving, access

characteristics, and satisfaction (Chen, Yen, & Hwang, 2012; Kurilovas & Juskeviciene, 2015; Sadaf, Newby, & Ertmer, 2012).

1.4. Web 2.0 Tools and Teacher Education

The International Society for Technology in Education (ISTE, 2016) has embarked on digital citizens to be raised that can use internet and digital tools effectively and appropriately in order for getting prepared for the uncertain future conditions. Similarly, Kumar & Vigil (2011) have emphasized that 21st century teachers are to produce digital content and integrate it with K-12 students in learning activities. Therefore, teachers are expected to have digital qualifications in order to apply Web 2.0 tools effectively in their classrooms (Sadaf, Newby, & Ertmer, 2012; Solomon & Schrum, 2007).

Kumar & Virgil (2011) have come into conclusion that present prospective teachers with high digital skills will not be able to transfer these skills into their future classrooms. One of the reasons presented in the same study is that prospective teachers are not individually able to bridge technology, pedagogy, and content knowledge. Additionally, similar research findings have indicated that prospective teachers are not well trained to apply Web 2.0 tools for effective instructional purposes (Lei, 2009; Kumar & Vigil, 2011). These results show that there is a need for frameworks in which technology, pedagogy, and content knowledge can be integrated.

In order to equip prospective teachers with digital skills, it is widely known and accepted that Technological Pedagogical and Content Knowledge (TPACK) competencies are more than necessary. Studies regarding TPACK competencies for present and future world have been on the front burner so far (Harris & Hofer, 2011; Koehler & Mishra, 2005). These studies have indicated that teacher-training programs should not be separated from TPACK skills and the present students in these programs have been widely surveyed for the issue. The common view states that TPACK is one of the expected and required competency for a teacher-to-be. However, different from common technologies, Web 2.0 tools have different technologies to be known. There needs to be different concentration for Web 2.0 tools to be actively used and benefited. Therefore, Web Pedagogical Content Knowledge (W-PCK) framework that was built upon PCK (Shulman, 1986) and TPACK (Mishra & Koehler, 2006) by Lee & Tsai (2010) to effectively use Web 2.0 tools specifically.

W-P CK framework consists of three knowledge fields; Content (C), Pedagogy (P) and Web (W). According to Lee & Tsai (2010), Web knowledge covers the use of both related tools and the their functions for communication or interaction. Content knowledge covers relevant subject area. Pedagogical knowledge covers teaching and learning processes and practices or methods applied in the processes. As seen in Figure 1, the W-PCK framework consists of four components, resulting in interaction between Content, Pedagogy and Web. Pedagogical Content Knowledge (PCK) contains pedagogical knowledge appropriate for the content. Web Content Knowledge (WCK) includes the use of Web tools that are specific to the subject area. Web Pedagogy Knowledge (WPK), includes the use of Web tools in pedagogical patterns. Web Pedagogical Content Knowledge (WPCK) involves the use of appropriate pedagogical patterns and Web tools in the teaching of the subject area.
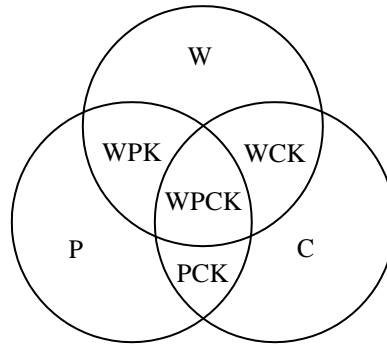
Figure 1. Web Pedagogical Content Knowledge (Lee & Tsai, 2010)

There are studies (Akgün, 2013; Arabacıoğlu & Dursun, 2015; Gömleksiz & Fidan, 2011; Kavanoz, Yüksel, & Özcan, 2015) basically describing pre-service teachers' W-PCK and their differences regarding age, gender, department, and additionally regarding self-efficacy. For a prospective teacher, it is important to have W-PCK competencies during training. Additionally, teacher training programs concern microteachings sincerely. They are one of the most effective methods in teacher training to foster prospective teachers to practice what they have learned so far and present their teaching skills. During the integration of Web 2.0 tools application in microteachings, a question appears: Is it possible to upskill prospective teachers both with teaching confidence and effective use of Web 2.0 tools in microteaching? The answer is not known for sure yet. With this perspective, this study investigates whether Web 2.0 tools applied in microteachings have an effect on prospective teachers W-PCK and teacher self-efficacy. Therefore, the questions below were tried to be answered:

- Does applying Web 2.0 tools in microteaching affect prospective teachers' W-PCK self-efficacies?
- Does applying Web 2.0 tools in microteaching affect prospective teachers' teacher self-efficacies?
- Is there any relationship between prospective teachers' W-PCK and teacher self-efficacies?
- What are the opinions of prospective teachers on applying Web 2.0 tools in microteaching?

2. Method

2.1. Research Design

The study was conducted in converging design of mixed method. As Creswell & Plano Clark (2011) stated, in this method, the qualitative and quantitative data were collected and analyzed at the same time. The equal priority of both data were gathered in interpretation section. Therefore, the quantitative data and qualitative data results supported one another.

The quantitative part of the study was conducted on single group pre-test/post-test quasi-experimental design. The independent variable of the study was effective Web 2.0 tools implementation in microteaching. The dependent variables of the study were prospective teachers' W-PCK and teacher self-efficacy. The scores of W-PCK self-efficacy were gathered through the "Web Pedagogical Content Knowledge Scale" and of teacher self-efficacy were gathered through the "Teacher Self-Efficacy Scale". After the seven-week

experimental process, the qualitative part of the research, were gathered through open ended question form consisting of prospective teachers' opinions on Web 2.0 tools. The design of the study was presented in Table 1.

Table 1. *Design of the study*

N	Pre-test	Treatment (7 weeks)	Post-test
24	W-PCK & Teacher self-efficacy	T	W-PCK & Teacher self-efficacy Open-ended question form

2.2. Participants

The study was conducted with 24 Tourism and Hotel Management Department students who participated in teacher training certificate program in Faculty of Education in a state university in Turkey. Of the 24 students, there were 10 females and 14 males. Participants' ages ranged from 22 to 30 years. The convenience sampling method was used in selecting the participants. In convenience sampling that is commonly used in both quantitative and qualitative studies, participants are selected because of their accessibility or convenience (Yıldırım & Şimşek, 2013). The participants received the treatment during course weeks and there did not exist any compelling force of the instructor on the participants due to the nature of the academic flow. They were also informed about the study beforehand and they voluntarily participated.

2.3. Context and Process

During the course maintenance, the instructor as the first researcher followed a sequence for an instructional design that is commonly known as ADDIE (analyze, design, develop, implement, evaluate) model. The instructor directed prospective teachers to follow this sequence while deciding their materials and matching them with purposes. The course was completed in 7 weeks in total, with 8 hours in each week. While selecting Web 2.0 tools, microteaching procedure and the application appropriateness were taken into consideration. In order for preservice teachers to prepare effective and attractive presentations with microteachings, Prezi; to prepare concept maps, Bubbl.us; to involve all students in evaluation process and determine learning gaps, game based Kahoot! and Puzzlemaker; and to share all the content, Edmodo tools were selected. The content of the course is presented in Table 2.

Table 2. *Content of the course*

	Duration	Content	Outcomes
Week 1	4	Introduction, pre-tests Identify instructional goals Task-job-goal analysis	Instructional goals Content Analysis report
	4	Learner analysis Instructional analysis	
Week 2	4	Writing performance objectives Developing assessment instruments	Instructional objectives Assessment instruments
	4	Developing instructional strategies Developing lesson plans	Lesson plans Design report
Week 3	4	Using Web 2.0 tools	Instructional materials

	4	Using Web 2.0 tools and developing instructional materials			
Week 4	4	Using Web 2.0 tools and developing instructional materials	Instructional materials		
	4	Microteaching	Feedback and revising instruction		
Week 5	4	Microteaching	Feedback and revising instruction		
Week 6 **Week 7**	4	Microteaching	Feedback and revising instruction		

2.4. Data Collection Tools

For the quantitative data collection, "Web Pedagogical Content Knowledge (W-PCK) Scale" and "Teacher Self-Efficacy Scale" were applied. W-PCK scale, which was developed by Lee & Tsai (2010), and adopted into Turkish by Horzum (2011), has five sub-dimensions namely *web-general, web-communication, web-content knowledge, web-pedagogical content knowledge, and attitudes towards web-based instruction*. The Cronbach alpha level of the scale was determined as .94. In this study, the alpha level was found as .91. The "Teacher Self-Efficacy Scale" which was developed by Tschannen-Moran & Hoy (2001) and adopted into Turkish by Çapa, Çakıroğlu, & Sarıkaya (2005) has three factors namely *increasing student participation, using instructional strategies, and classroom management*. The Cronbach alpha level of the scale was determined as .93. In this study, the alpha level of the scale was found as .96.

For the qualitative data collection, a structured form including four open ended questions were used. The form was developed by the researchers regarding related literature (Daher & Lazarevic, 2014; Davis, 1989; Sadaf, Newby, & Ertmer, 2012), in which participants can evaluate the course content, applied Web 2.0 tools, their personal and professional developments, and the researchers' professional experiences. After questions were formed, they were sent to three experts, one from Educational Sciences and two from Computer Education and Instructional Technologies, for their opinions. Regarding the feedback from experts, the form was revised and finalized for participants' responses. In the form, following questions were asked: *Have you ever had any difficulties while using Web 2.0 tools? If yes, with which tool? Was the course effective to provide you with teaching skills? Please explain how. What are your likes and dislikes about the course? Please explain why. Which Web 2.0 tools that you learnt during the course would you like to use primarily in your future classroom? Please explain why.*

2.5. Data Analysis

In the analysis of the quantitative data, to determine the difference between pre-test and post-test, paired samples t-test was utilized. Before analyses, normality of difference between pre and post-test scores of W-PCK and teacher self-efficacy scales were tested. According to the Shapiro–Wilk test, scores for difference between pre and post-test scores of W-PCK ($p=.14$) and teacher self-efficacy ($p=.11$) were distributed normally. In many studies (Alvarez, Brown, & Nussbaum, 2011; Laru, Näykki, & Järvelä, 2012; Lehiste, 2015), it has been observed that paired samples t-test, which is a parametric test, has been applied for small groups of 17 and 20 participants. Therefore, paired samples t-test was used to examine the differences between the pre-tests and the post-tests in this study. The SPSS 22.0 program was used for the data analysis. Regarding the significance level, the value of .05 was

considered. For the calculation of the effect size, the $\eta^2 = t^2 / (t^2 + (n - 1))$ equation was used (Büyüköztürk, 2015), while the following criteria for the values of η^2 (.01 small effect, .06 medium effect, .14 large effect) were considered for the effect size (Büyüköztürk, 2015). The post-test scores of W-PCK and Teacher Self-Efficacy scale were proved that the scores were normally distributed regarding Shapiro-Wilk test results found respectively as .29 and .06. Therefore, the relationship between scores of both scales were examined applying Pearson correlational analysis.

For the qualitative data analysis, content analysis was conducted (Miles & Huberman, 1994). In content analysis, it is aimed to combine similar data under themes and to present them so that readers can understand (Yıldırım & Şimşek, 2013). The induction approach was adopted when data analysis was conducted. Codes were generated by direct citation of the responses to the questions, and these codes were collected under certain themes. While presenting the data, only the themes were used to reflect the views of the participants. Frequencies were also indicated. The two researchers have conducted the coding process independently in creating codes and forming the themes in order to provide reliability. Inter-coder reliability was set through percentage agreement (95%). For further validity and reliability of the results in the research, the data were presented in details and direct quotations were witnesses as much as possible (Yıldırım & Şimşek, 2013).

3. Findings

3.1. Results of W-PCK Scale Scores

In order to examine the difference between pretest-posttest scores for the W-PCK scale, paired samples t-test was conducted. The t-test results were presented in Table 3. Based on t-test results, it was found that there was a significant difference between participants' pre-posttests scores. The prospective teachers as participants who took course with Web 2.0 tools for the instructional purposes showed increase in their scores. The results of participants according to W-PCK scale's sub dimension were found as for web-general t(23)=3.43, p<.05; web-communication t(23)=2.51, p<.05; web-content knowledge t(23)=3.51, p<.05; web-pedagogical content knowledge t(23)=5.44, p<.05; and attitudes towards web-based instruction t(23)=7.16, p<.05; and as for the total scale scores t(23)=7.16, p<.05. The effect sizes of web-general (.34), web-communication (.22), web-content knowledge (.35), attitudes towards web-based instruction (.41), web-pedagogical content knowledge (.56), and total scale scores (.69) were found as large level. These results show that the course with its implementations have had an important effect size on participants' W-PCK. Additionally, it can be assumed that prospective teacher might get familiar with the course content and the technology-integrated instructions therefore, their W-PCK scores might increase.

Table 3. *T-test results of W-PCK Scale*

	Measurement	N	X	S	df	*t*	*p*	η^2
Web-General	Pretest	24	32.50	2.73	23	3.43	*0.00*	0.34
	Posttest	24	33.96	1.71				
Web-Communication	Pretest	24	18.21	2.38	23	2.51	*0.02*	0.22
	Posttest	24	19.21	1.53				
Web-Content Knowledge	Pretest	24	22.29	2.29	23	3.51	*0.00*	0.35
	Posttest	24	24.00	1.79				
Web-Pedagogical	Pretest	24	33.63	4.26	23	5.44	*0.00*	0.56

Content Knowledge	Posttest	24	37.63	2.65				
Attitudes towards	Pretest	24	25.75	3.35	23	3.99	*0.00*	0.41
Web-based	Posttest	24	29.67	5.04				
Instruction								
Total Scale	Pretest	24	132.38	11.02	23	7.16	*0.00*	0.69
	Posttest	24	144.46	9.17				

3.2. Results of Teacher Self-Efficacy Scale Scores

In order to examine the difference between pretest-posttest scores for the self-efficacy scale, paired samples t-test was conducted. The t-test results were presented in Table 4. The results showed that there was a significant difference between participants' pretest-posttest scores. The implementations were significant for the ensuring students participation $t(23)=4.43$, $p<.05$, applying instructional strategies $t(23)=5.23$, $p<.05$ and classroom management $t(23)=3.72$, $p<.05$ sub dimensions and for the total scale scores $t(23)=4.91$, $p<.05$. The effect sizes for ensuring students' participation (.46), classroom management (.37), applying instructional strategies (.54), and total scale (.51) were found as large level. These results show that the course with its implementations have had an important effect size on participants' teacher self-efficacy.

Table 4. *T-test results of Teacher Self-Efficacy Scale*

	Measurement	N	X	S	df	*t*	*p*	η^2
Ensuring Students	Pretest	24	52.96	2.73	23	4.43	*0.00*	0.46
Participation	Posttest	24	59.46	1.71				
Applying	Pretest	24	53.08	2.38	23	5.23	*0.02*	0.54
Instructional	Posttest	24	61.50	1.53				
Strategies								
Classroom	Pretest	24	55.08	2.29	23	3.72	*0.00*	0.37
Management	Posttest	24	60.58	1.79				
Total Scale	Pretest	24	161.13	4.26	23	4.91	*0.00*	0.51
	Posttest	24	181.54	2.65				

3.3. The Relation between Prospective Teachers' W-PCK and Self-Efficacy Perceptions

Descriptive findings and correlation coefficient for the W-PCK and teacher self-efficacy post-test scores of the prospective teachers were presented in Table 5. As the result of the examination, it was found that there was a medium level positive relation ($r=.502$, $p<.05$) between W-PCK and teacher self-efficacy scores. It can be interpreted as that W-PCK self-efficacy increase; teacher self-efficacy perceptions increase as well. When the determination coefficient ($r^2=0.25$) was taken into consideration, it can be commented that 25% of the total variance in prospective teachers' self-efficacies can be explained with W-PCK.

Table 5. *W-PCK and Teacher Self-Efficacy Scales' descriptive results and correlational coefficients*

Variables	X	S	1	2
1- TAPCK-W	144.46	9.17	1	
2- Teacher self-efficacy	181.54	21.50	.502[*]	1

$p<.05$, n=24

3.4. Prospective Teachers' Views On

3.4.1. Web 2.0 tools

Prospective teachers mostly responded positively for the Web 2.0 tools and their application level. Most of them responded as they did not have any difficulty in applying tools, and the rest responded as they had minor problems in applying the tools in the beginning, but then adapted quickly. The themes were determined as *no difficulty in applying tools* and *minor difficulties at the first application*. The sample responses were presented in Table 6. below.

Table 6. *Themes and sample responses on web 2.0 tools*

Themes	Sample Responses
No difficulty in applying tools (10)	*"All the applications were easy and simple."* *"I did not have any difficulty."* *"All the applications were clear and easy to understand."*
Minor difficulties at the first application (10)	*"When I first used the application, I had some minor problems. Then I realized how enjoyable the tools were and I started to use without any problem."* *"When I first used Prezi, I found it hard. Then I got used and quickly applied."*

3.4.2. The course

Prospective teachers reflected positive attitudes towards course period they had experienced. They mostly focused on their developed skills and competencies for technology integration in instruction and professional confidence. These responses were formed under four themes namely, *developed technology use skills, teaching skills, active participation in course period*, and *review chances for incompetency*. Sample responses for the related themes were presented in Table 7.

Table 7. *Themes and sample responses on the course*

Themes	Sample Responses
Developed technology use skills (16)	*"We have learned many web tools and how to apply these tools in instruction."* *"It helped us to decide which web tools we could apply and how to do it."*
Developed teaching skills (12)	*"I have had many presentations but they were only Power Point slides. The tools I learned through this course contributed to my presentation skills a lot. I am sure that my teaching life will continue with joy and my students will have fun in courses."* *"Thanks to practices we performed we overcome our anxiety and had experiences."* *"We have learned instructional flow and how to manage the instructional time thanks to microteachings."*
Active participation in course period (7)	*"It was a complete student friendly course so that we all had great time while learning."* *"While learning and using web tools, the whole class actively involved."*
Review chances for incompetency (2)	*"The feedbacks when we made a mistake helped us to review and correct ourselves."*

3.4.3. The course likes and dislikes

Prospective teachers stated that they enjoyed the fact that they were actively participating, finding opportunities to teach, using Web 2.0 tools, seeing the instructor as a model, earning self-confidence, and achieving permanent learning thanks to the course they took. The themes were determined as *active participation, lecturing, using web 2.0 tools, instructor as a model, higher self-confidence,* and *realization of permanent learning*. Prospective teachers stated that they liked the procedure of the course. The sample responses with related themes were presented in Table 8.

Prospective teachers only stated that they did not like the duration of the lecture as a dislike. The responses were recorded as, *"The only thing I did not like is that the presentations' duration were very short.", "The given time for the presentations were too short.", "Since microteaching duration was very short, we rushed."*

Table 8. *Themes and sample responses on course likes*

Themes	Sample Responses
Active participation (4)	*"The fact that we actively participated in the course and learned many new things were what I liked most."* *"It was a very effective experience to actively involved in the course with microteachings."*
Lecturing (4)	*"To prepare for the lesson individually, to come here and teach like a teacher was very productive."* *"For the first time I taught in a planned way without memorization. This is what I liked."*

Using web 2.0 tools (4)	*"I have learned many web tools that I liked and they will help me in the future."* *"This was a very productive period I think I have learned to use the technology more usefully."*
Instructor as a model (3)	*"It is a fact that the instructor practiced everything he taught in his own lesson was very effective."* *"I liked the fact that the instructor's being model for us about the use of technology."* *"I would like to apply my own model of this lesson to my students in order to achieve high levels of achievement in the same way when I become a teacher."*
Higher self-confidence (2)	*"I liked that my self-confidence increased in the course."* *"My self-confidence developed. That's what I liked."*
Realization of permanent learning (1)	*"I am delighted that I still have not forgotten what I learned."*

3.4.4. The web 2.0 tools for future classrooms

Prospective teachers were found very willing to use web tools in their future classrooms. They mostly stated that they were tend to use Kahoot and Prezi, which were of the introduced tools in the course. They responded positively for Kahoot *(14)*, Prezi *(13)*, Bubbl.us *(8)*, and Edmodo *(7)* applications as, *"Kahoot is very enjoyable and informative.", "With Kahoot, the courses can be more fun and meaningful through using post exercises at the end of class hours.", "I will use Prezi to present attentive visual presentations for my students.", "I plan to use Prezi in presentation section of the course.", "With Bubbl.us, I can use concept maps since it helps to present the issue simply for the students.", "To form concept maps is easy and saves time.", "For the communication with students, I will apply Edmodo.", "To share materials, homework, and exercises, I will apply Edmodo."*

4. Results and Discussion

Regarding the study results, it was noticed that the microteachings and the applied Web 2.0 tools during the course procedure have changed both prospective teachers' W-PCK and Self-Efficacy scale scores significantly. The effect sizes on both W-PCK and Self-Efficacy scores were found important to report that implementations during the course procedure were effective. The medium level relation between W-PCK and Self-Efficacy results also showed that their scores increased simultaneously, which explains that the more prospective teachers engage in acquaintance and competence with web tools the more they increase in self-efficacy.

As observed in findings above, effective use of Web 2.0 tools in microteaching affected prospective teachers' W-PCK and self-efficacy. Additionally, their own participation and motivation were positively influenced by these tools by means of applying them individually in microteachings. The scale scores and responses showed that Web 2.0 tools broadened new horizons for future teachers.

Additionally, the views of prospective teachers on Web 2.0 tools, the course, likes and dislikes, and plans for future classrooms presented further implications for the implementations. The views on Web 2.0 tools have presented that adaptation in technology for instructional purposes may not be so difficult to manage. Therefore, the prejudices such as fear of loss of time to figure out on knowing, selecting, and adapting the right Web 2.0 tool for instructional purposes can be eliminated before future teachers start their own

implementations. The views on the course in general have presented that prospective teachers have had the opportunity for developing technology skills and therefore teaching skills at the same time, which also have resulted in active participation and immediate feedbacks. Followingly, prospective teachers commented on the course procedure by evaluating it in different ways. They have mentioned that the course itself with implementations created a joyful and informative atmosphere. With the microteachings conducted during the course with the modelling of the instructor have made what has been learnt more permanent and meaningful. The prospective teachers in the study also shared their opinions on future classroom implementations by referring the Web 2.0 tools that were used in the course. Therefore, it is likely to say that they are willing to apply similar or same web tools in their future classrooms effectively.

First of all, accepting that Web 2.0 tools are pedagogically meaningful, prospective teachers reflected high tendency to transfer these tools into their real classroom teachings. However, it can be said that modelling teachers with such instructional technologies is also of great importance. Utami & Nafi'ah (2016) stress in their study that prospective teachers as observers in a course flow can gain many experiences from microteachings and their being modelled. Therefore, in pre-service programs, instructors should be encouraged to stimulate purposeful technologies regarding departments, course subjects, classroom dynamics and many other teacher/student based variables. Furthermore, upskilling prospective teachers with current competencies and frequently used either personally or collaboratively Web 2.0 tools influenced their self-efficacy as observed in similar studies (Curts et al., 2008; Niederhauser & Perkmen, 2008). In addition, Kavanoz, Yüksel, & Özcan (2015) and Akgun (2013) have stated that there is a positive correlation between web pedagogical knowledge and self-efficacy of pre-service teachers as similar to this study's findings. In this study, it was found that the prospective teachers had a meaningful relationship between W-PCK and teacher self-efficacy. This result indicated that 25% of the total variance in prospective teachers' self-efficacies can be explained with W-PCK also. It shows that teacher self-efficacy can be increased by enhancing W-PCK competencies of teachers. As mentioned, Web 2.0 tools can be applied to enhance this bound of pre-service teachers for their further competencies in technology and instruction.

The study also emphasized that monotonous presentation of Web 2.0 tools by an instructor/trainer might not solely be enough to advance teaching skills. It is also necessary to give opportunity for pre-service teachers to try to design their own teaching flow with presented tools appropriate for their purposes in a theoretical framework. Mannathoko (2013) embarks that insufficient assistance and implementations before teaching practice in real school environment negatively affects school practice of prospective teachers who need more practice before the real start. Therefore, the hot crisis "the gap between theory and practice" can be closed smoothly. That is why microteachings should not be disregarded if the issue is teacher training.

Many researchers have reported that teachers' attitudes towards technology use are highly affected by the perceived benefits of technology and ease of use (Davis, 1989; Sadaf, Newby & Ertmer, 2012; Teo, Lee & Chai, 2008). In other words, when teachers develop a perception that the new technology offered to them facilitates access to instructional objectives and is easy to use, it affects their attitudes towards technology, and following their attitudes affect technology use behaviors (Davis, 1989). The results of the research on prospective teachers' views on Web 2.0 tools reveal that prospective teachers have a perception that Web 2.0 tools are easy to use and pedagogically meaningful. Positive perceptions, when the attitude sub-dimension of the W-PCK scale is evaluated together, have turned into positive attitude. This

indicates that prospective teachers can use Web 2.0 tools in their classes. Prospective teachers' views also show that they are willing to use Web 2.0 tools in their future classes.

The study results in general highlighted that technology integration is more effective when it is profoundly designed and introduced. Requiring from teachers to be active users of any technology agent in classroom environment may not be convincing and operative all the time since it depends on many different dynamics. In order for teachers to actively and effectively use of technology tools in different educational purposes there needs an intersection of knowing the technology tools that will be applied, transferring them into content, and relating them with a meaningful outcome and learning (Angeli & Valanies, 2009). That is, after introducing teachers with any technology tool and presenting the reason, instructional purpose and practice opportunity tend to be more attentive and facilitative. Therefore, as it was implemented in the study, it is better to inform and train pre-service teachers and/or in-service teachers not with a random technology-based tool, but instead with goal-oriented technology tools.

Technology integration in education is both complicated and difficult as a process. There are first and second obstacles to overcome (Ertmer, 1999, 2005; Hew & Brush, 2007). A third type of obstacle is teacher education (Tsai & Chai, 2012). One of the most effective methods used for teacher education is microteaching method. It is inevitable that the microteaching method should be updated to include the use of modern technologies. These research results, which focus on the use of Web 2.0 tools in the microteaching methodology, have showed that the correct use of Web 2.0 tools in microteaching has significantly increased the W-PCK and teacher self-efficacy of prospective teachers. Correct use has the following steps: a combination of an instructional design model such as the ADDIE model and the W-PCK framework; Instructor's modeling for prospective teachers by thinking aloud how Web 2.0 tools are used in what stage, for what purpose and how; Prospective teachers have their own stages of making their own teaching designs.

As final words, the study has answered that it is possible to affect prospective teacher' W-PCK and teacher self-efficacy by using Web 2.0 tools and practicing them in microteachings. However, the point proposed by the study is that describing and presenting these tools may not be a reinforce for prospective teachers who are supposed to be the users and transponders of these tools. It has been seen that prospective teachers will tend to more eager and curious to apply these tools when they are explained which tools is more appropriate for which part of an instructional period and how they can use it for an effective teaching.

The results of the research provide important implications for decision makers and educators. In the case of correct use of Web 2.0 tools in micro-education, W-PCK and teacher self-efficacy of prospective teachers can develop, prospective teachers can develop positive attitude towards Web 2.0 tools and may become willing to use these technologies in their class in the future. The applied education can contribute to the teachers training in terms both of their self-confident and of technology use. For this reason, curricula used in teacher education can be updated to include microteaching practices that allow the use of Web 2.0 tools.

References

Alvarez, C., Brown, C., & Nussbaum, M. (2011). Comparative study of netbooks and tablet PCs for fostering face-to-face collaborative learning. *Computers in Human Behavior, 27*(2), 834–844. doi:10.1016/j.chb.2010.11.008

Akcaoglu, M., Gumus, S., Bellibas, M. S., & Boyer, D. M. (2015). Policy, practice, and reality: Exploring a nation-wide technology implementation in Turkish schools. *Technology, Pedagogy and Education, 24*(4), 477–491. doi:10.1080/1475939X.2014.899264

Akgün, F. (2013). Öğretmen adaylarının web pedagojik içerik bilgileri ve öğretmen öz-yeterlik algıları ile ilişkisi [Preservice teachers' web pedagogical content knowledge and relationship between teachers' perceptions of self-efficacy]. *Trakya University Journal of Education, 3*(1), 48–58.

Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT-TPCK: Advances in technology and pedagogical content knowledge (TPCK). *Computers and Education, 52*, 154-168.

Arabacıoğlu, T., & Dursun, F. (2015). Öğretmen adaylarının web pedagojik içerik bilgisi algı düzeylerinin incelenmesi [Examination of the web pedagogical content knowledge of pre-service teachers in terms of some variables]. *Kastamonu Education Journal, 23*(1), 197–210.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215. doi:10.1037/0033-295X.84.2.191

Bandura, A. (1986). The explanatory and predictive scope of self-efficacy theory. *Journal of Social and Clinical Psychology, 4*(3), 359–373. doi:10.1521/jscp.1986.4.3.359

Brown, S. A. (2012). Seeing Web 2.0 in context: A study of academic perceptions. *Internet and Higher Education, 15*(1), 50–57.

Büyüköztürk, Ş. (2015). *Sosyal bilimler için veri analizi el kitabı* (21. bs.) [Data analysis handbook for social sciences] (21st ed.)]. Ankara: Pegem Academy Publishing.

Cavin, R. (2008). Developing technological pedagogical content knowledge in preservice teachers through microteaching lesson study. *Technology and Teacher Education Annual. 19*(8), 5214.

Chen, S. C., Yen, D. C., & Hwang, M. I. (2012). Factors influencing the continuance intention to the usage of Web 2.0: An empirical study. *Computers in Human Behavior, 28*(3), 933–941. doi:10.1016/j.chb.2011.12.014

Churchill, D. (2011). Web 2.0 in education: A study of the explorative use of blogs with a postgraduate class. *Innovations in Education and Teaching International, 48*(2), 149–158.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research.* Los Angeles, CA: Sage.

Curts, J., Tanguma, J., & Peña, C. M. (2008). Predictors of Hispanic school teachers' self-efficacy in the pedagogical uses of technology. *Computers in the Schools, 25*(1–2), 48–63. doi:10.1080/07380560802157766

Çapa, Y., Çakıroğlu, J., & Sarıkaya, H. (2005). The development and validation of a Turkish version of teachers' sense of efficacy scale. *Education and Science, 30*(137), 74–81.

Çoban, A. (2015). Öğretmen eğitiminde mikro-öğretim ve farklı yaklaşımlar [Micro-teaching and different approaches in teacher training]. *Electronic Journal of Social Sciences, 14*(53), 219–231.

Daher, T., & Lazarevic, B. (2014). Emerging instructional technologies: Exploring the extent of faculty use of web 2.0 tools at a midwestern community college. *TechTrends: Linking Research & Practice to Improve Learning, 58*(6), 42–50. doi:10.1007/s11528-014-0802-1

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance. *MIS Quarterly, 13*(3), 319–339. doi:10.2307/249008

Ertmer, P. A. (1999). Addressing first- and second-order barriers to change: Strategies for technology integration. *Educational Technology Research and Development, 47*(4), 47–61. doi:10.1007/BF02299597

Ertmer, P. A. (2005). Teacher pedagogical beliefs: The final frontier in our quest for technology integration? *Educational Technology Research and Development, 53*(4), 25–39. doi:10.1007/BF02504683

Ertmer, P. A., & Ottenbreit-Leftwich, A. T. (2010). Teacher technology change: How knowledge, confidence, beliefs, and culture intersect. *Journal of Research on Technology in Education, 42*(3), 255–284. doi:10.1080/15391523.2010.10782551

Gömleksiz, M. N., & Fidan, E. K. (2011). Self-efficacy perception levels of prospective teachers' enrolled at pedagogical formation course toward web pedagogical content knowledge. *Journal of Turkish Studies, 6*(4), 593–620. doi:10.7827/TurkishStudies.2910

Görgen, İ. (2003). Mikroöğretim uygulamasının öğretmen adaylarının sınıfta ders anlatımına ilişkin görüşleri üzerine etkisi [The effect of microteaching practices on student teachers' views of giving lessons in the classroom]. *Hacettepe University Journal of Education, 24*, 56–63.

Harris, J. B., & Hofer, M. J. (2011). Technological pedagogical content knowledge in action: A descriptive study of secondary teachers' curriculum-based, technology-related instructional planning. *Journal of Research on Technology in Education, 43*(4), 211.

Hartshorne, R., & Ajjan, H. (2009). Examining student decisions to adopt Web 2.0 technologies: Theory and empirical tests. *Journal of Computing in Higher Education, 21*(3), 183–198.

Hew, K. F., & Brush, T. (2007). Integrating technology into K-12 teaching and learning: current knowledge gaps and recommendations for future research. *Educational Technology Research & Development, 55*(3), 223–252. doi:10.1007/s11423-006-9022-5

Hew, K. F., & Cheung, W. S. (2013). Use of Web 2.0 technologies in K-12 and higher education: The search for evidence-based practice. *Educational Research Review, 9*, 47–64.

Horzum, M. B. (2011). Web Pedagojik içerik bilgisi ölçeği'nin Türkçeye uyarlaması [Adaptation of web pedagogical content knowledge survey to Turkish]. *Elementary Education Online, 10*(1), 257–272.

ISTE (2016). Redefining learning in a technology-driven world. Retrieved 15 September 2017, from https://www.iste.org/standards/for-educators.

Kavanoz, S., Yüksel, H. G., & Özcan, E. (2015). Pre-service teachers' self-efficacy perceptions on web pedagogical content knowledge. *Computers & Education, 85*, 94–101. doi:10.1016/j.compedu.2015.02.005

Koehler, M. J., & Mishra, P. (2005). Teachers learning technology by design. *Journal of Computing in Teacher Education, 21*(3), 94–102. doi:10.1080/10402454.2005.10784518

Kumar, S., & Vigil, K. (2011). The net generation as preservice teachers: Transferring familiarity with new technologies to educational environments. *Journal of Digital Learning in Teacher Education, 27*(4), 144-153.

Kurilovas, E., & Juskeviciene, A. (2015). Creation of Web 2.0 tools ontology to improve learning. *Computers in Human Behavior, 51*, 1380–1386. doi:10.1016/j.chb.2014.10.026

Laru, J., Näykki, P., & Järvelä, S. (2012). Supporting small-group learning using multiple Web 2.0 tools: A case study in the higher education context. *The Internet and Higher Education, 15*(1), 29–38. doi:10.1016/j.iheduc.2011.08.004

Lee, M. H., & Tsai, C. C. (2010). Exploring teachers' perceived self-efficacy and technological pedagogical content knowledge with respect to educational use of the World Wide Web. *Instructional Science, 38*(1), 1–21. doi:10.1007/s11251-008-9075-4

Lehiste, P. (2015). The impact of a professional development program on in-service teacher's TPACK: A study from Estonia. *Problems of Education in the 21st Century, 66*, 18–28.

Lei, J. (2009). Digital natives as preservice teachers: what technology preparation is needed? Journal of Computing in Teacher Education, 25(3), 87–97.

Mannothoko, M. C. (2013). Does teaching practice effectively prepare student-teachers to teach creative and performing arts? The case of Botswana. *International Journal of Higher Education, 2*(2), 115-121. doi:10.5430/ijhe.v2n2p115

Meyer, K. A. (2010). A comparison of Web 2.0 tools in a doctoral course. *Internet and Higher Education, 13*(4), 226–232. doi:10.1016/j.iheduc.2010.02.002

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). Thousand Oaks, CA: Sage.

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054. doi:10.1111/j.1467-9620.2006.00684.x

Niederhauser, D. S., & Perkmen, S. (2008). Validation of the intrapersonal technology integration scale: Assessing the influence of intrapersonal factors that influence technology integration. *Computers in the Schools, 25*(1–2), 98–111. doi:10.1080/07380560802157956

O'Reilly, T. (2007). What is web 2.0: Design patterns and business models for the next generation of software. Retrieved 5 September 2017, from https://mpra.ub.uni-muenchen.de/4578/1/MPRA_paper_4578.pdf

Sadaf, A., Newby, T. J., & Ertmer, P. A. (2012). Exploring pre-service teachers' beliefs about using Web 2.0 technologies in K-12 classroom. *Computers & Education, 59*, 937–945. doi:10.1016/j.compedu.2012.04.001

Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14. doi:10.3102/0013189X015002004

Solomon, G., & Schrum, L. (2007). *Web 2.0: New tools, new schools*. Eugene, OR: International Society for Technology in Education.

Teo, T., Lee, C. B., & Chai, C. S. (2008). Understanding preservice teachers' computer attitudes: Applying and extending the Technology Acceptance Model. *Journal of Computer Assisted Learning, 24*(2), 128−143

Tondeur, J., van Braak, J., Sang, G., Voogt, J., Fisser, P., & Ottenbreit-Leftwich, A. (2012). Preparing pre-service teachers to integrate technology in education: A synthesis of qualitative evidence. *Computers & Education, 59*(1), 134–144. doi:10.1016/j.compedu.2011.10.009

Tsai, C. C., & Chai, C. S. (2012). The "third"-order barrier for technology-integration instruction: Implications for teacher education. *Australasian Journal of Educational Technology, 28*(6), 1057–1060.

Tschannen-Moran, M., & Hoy, A. W. (2001). Teacher efficacy: capturing an elusive construct. *Teaching and Teacher Education, 17*(7), 783–805. doi:10.1016/S0742-051X(01)00036-1

Utami, I. W. P., & Nafi'ah, U. (2016). A model of microteaching lesson study implementation in the prospective history teacher education. *Journal of Education and Practice, 7*(27), 10-14.

Yıldırım, A., & Şimşek, H. (2013). *Sosyal bilimlerde nitel araştırma yöntemleri* (9. bs.) [Qualitative research methods in the social sciences (9th ed.)]. Ankara: Seçkin Publishing.

THE EFFECT OF PROBLEM BASED LEARNING MODEL ON STUDENT MATHEMATICS LEARNING OUTCOMES VIEWED FROM CRITICAL THINKING SKILLS

Heri Mulyanto (iD)

Sebelas Maret University, Surakarta, Indonesia

eliezerherimulyanto@yahoo.com

Gunarhadi (iD)

Sebelas Maret University, Surakarta, Indonesia

gunarhadi@fkip.uns.ac.id

Mintasih Indriayu (iD)

Sebelas Maret University, Surakarta, Indonesia

mientasihindriayu@yahoo.com

Biodata(s)

Heri Mulyanto is a Student Magister of Primary School Teacher Education Sebelas Maret University and he is a teacher of Kalam Kudus Elementary School in 5 grade.

Prof. Gunarhadi, MA, Phd. is a lecturer in The Faculty of Teacher Training and Education Sebelas Maret University Indonesia.

Dr. Mintasih Indriayu, M.Pd. is a lecturer in The Faculty of Teacher Training and Education Sebelas Maret University Indonesia.

THE EFFECT OF PROBLEM BASED LEARNING MODEL ON STUDENT MATHEMATICS LEARNING OUTCOMES VIEWED FROM CRITICAL THINKING SKILLS

Heri Mulyanto

eliezerherimulyanto@yahoo.com

Gunarhadi

gunarhadi@fkip.uns.ac.id

Mintasih Indriayu

mientasihindriayu@yahoo.com

Abstract

This study aims to determine the effect of applying problem based learning model on students' mathematics learning outcomes viewed from critical thinking skills of fifth grade students of the private primary school in the Surakarta region. This research is an experimental research using sample of 309 students consisting of experimental class of 153 students and control class of 156 students. Method of data analysis using Two-Way Anova. Results of this study indicate (1) There are significant differences in mathematics learning outcomes between students who following Problem Based Learning model with conventional model. (2) There is a significant difference in mathematics learning outcomes between students who have high and low critical thinking skills. (3) There is no interaction effect between learning model and critical thinking skills on mathematics learning outcomes.

Keywords: learning outcome, mathematics, learning, problem based learning, conventional, critical thinking skills

1. Introduction

Experimental research about effect interaction of learning model with critical thinking skill on learning achievement of mathematics of private primary school students in Surakarta is important. This research motivated by survey results of researcher, that results of science and mathematics learning achieved of fifth grade of the private primary school in Surakarta has not been maximized. Causes factor can be derived from the teacher factor that is learning model has not aroused the creativity of student thinking. Mathematics learning is still teacher-oriented so that students are less actively involved in learning, problem-solving and less critical thinking skills. Related with the problem, teachers need to develop innovative learning models such as Problem Based Learning (PBL).

According to Sanjaya (2007: 214) Problem Based Learning (PBL) is a series of learning activities that emphasize the process of solving problems faced scientifically. According to Glatthorn & Craft-Tripp (in Pecore, 2012) Problem Based Learning is a learning model based on constructivism theory that will make the students interest in learning and actively participate in the learning process. According to Muraray-Harvey, Pourshafie, and Reyes (2013) one way to create active learning is to apply problem-based learning models. Problem-based learning models emphasize broader PBM processes, creating opportunities to develop meaningful knowledge, skills and attitudes related to collaborative learning, so as to

build cooperative knowledge, help students to establish explicit relationships between attitudes toward cooperation and achieve learning outcomes, and identify specific collaborative skills required by students, and obtained through group collaboration. The theory developed by Glatthorn& Craft-Tripp (in Pecore, 2012), states Problem Based Learning is a model of learning based on constructivism theory that will make students interested in learning and active participation in the learning process. According to Muraray-Harvey, Pourshafie, and Reyes (2013) one way to create active learning is to apply problem-based learning models. The problem-based learning model emphasizes the broader process of teaching and learning, creating opportunities to develop meaningful knowledge, skills and attitudes related to collaborative learning, so as to build cooperative knowledge, help students to establish explicit relationships between attitudes toward cooperation and achieve learning outcomes; identify specific collaborative skills required by students, and obtained through group collaboration. The results of Frank (2011) suggest that a scientific approach in mathematics learning to improve skills and skills is therefore a strong constraint on basic education. In addition, the scientific approach can also improve competence in modern mathematics learning. Hmelo-Silver (in Lian, 2010: 24) states several strikes of PBL models such as a) Problem solving is a good technique to better understand the content of the lesson; b) Problem solving can improve students' ability and provide satisfaction to discover new knowledge by students; c) Problem solving can increase student activity during the learning process.

Mathematics learning outcome can also be influenced from student factors that is critical thinking skill. The theory proposed by Chaffe (2012: 4) states that critical thinking is the process of thinking to clarify one's understanding of something resulting in an intelligent decision. According to Moon (2008: 21) critical thinking is the ability to consider various information obtained from various different sources, then process them creatively and logically and analyze them to get the right conclusions. Previous research results of Karim &Normaya (2015) states that one of the primary goals of schooling is to form students' critical thinking skills and one of the subjects considered to teach critical thinking skills is mathematics. A child's skill to think critically well can shape behavior when dealing with problems with rational thought. Previous survey results in one of the private primary schools in Surakarta show that students only study mathematics drill, memorize concepts, theories, and test or test-oriented. Students are accustomed to receiving lessons from teacher and not trying to find or shape their own knowledge. Facione (2015) states, student's critical thinking skills can be measured with six dimensions of Interpretation, Analysis, Inference, Evaluation, Explanation, and Self-Regulation.

Previous research findings conducted by Ajai et al. (2013) states that there is a significant difference in post-test results of mathematics students based on learning models, post-test results of mathematics students who follow the PBL model learning is higher than conventional model. Previous research findings conducted by Fatade et al. (2013) states that there is a significant difference in post-mathematics test result of junior high school students in Nigeria based on learning model, post-test result of mathematics of students who follow PBL model learning is higher than control class students. The findings of previous research conducted by Fazriyah et al. (2017) stated that there is a significant difference of science learning outcomes of fifth grade students of Primary School Depok based on critical thinking skills. Students who have high critical thinking skills, students' science learning outcomes given an integrative thematic learning model is higher than in the fragmented learning model. Students who have low critical thinking skills, higher learning outcomes of natural science use a fragmented model. Previous research findings conducted by Erwin, Tellu and Kundera (2015) states there are significant differences in student learning outcomes in Biology lesson

in 4 senior high schoolofPalu in terms of students' critical thinking skills. Student learning outcomes that have high critical thinking skills is higher than students who have low critical thinking skills. The findings of this study support the results of previous research conducted by Nurwulandari, Muhardjito, and Mufti (2015) that there are significant differences in problem-solving abilities in terms of critical thinking skills in class X students in 1 senior high school Malang Year 2014/2015. Students with high critical thinking skills have higher problem-solving abilities than students with low critical thinking skills.

The results of a previous survey on one of the private primary schools in Surakarta showed that students only learn mathematics drill, memorize concepts, theories, and test or test-oriented. Students are accustomed to receiving lessons from what the teacher just says and not trying to find or shape their own knowledge. The problem motivates the researcher to analyze the influence of learning problem based learning model and critical thinking skills toward mathematics learning result and to analyze the influence of interaction between problem based learning model with critical thinking skills toward mathematics learning outcomes in V students of private primary school in Surakarta region.

Based on the problems and findings of previous research the researcher took the initiative to conduct research on the effect of Learning Problem Based Learning Model on Mathematics Learning Outcomes Viewed from the Critical Thinking Skills of Fifth grade Students of the Private Primary School in Surakarta Region Year 2016/2017.

2. Methods

This research is an experimental research using sample of 309 private primary school students in five sub-districts of Surakarta city in Year 2016/2017. Sample consists of experiment class are 153 students and control class are 156 students. The experimental class is a group of students who are subject to learning with a Problem Based Learning model (PBL) while the control class is a group of students who are subjected to conventional learning. Dependent variable in this research is mathematics learning outcomes. Dependent variables in this research are (1) learning PBL model and (2) critical thinking skills. Instrument of critical thinking skills in the form of questionnaires consists of 20 items of statements measured by six dimensions developed by Facione (2015) namely Interpretation, Analysis, Inference, Evaluation, Explanation, and Self-Regulation. This instrument presents two alternative answers Yes and No. The measurement scale for each odd numbered question gets a value of 5 for each item if it answers "Yes" ("No" = 0). For each even numbered question get a value of 5 for each item if it answers "No" ("Yes" = 0). The Cut of Point> 70 indicator is said to be critical. Measurement instrument of mathematics learning outcomes test aims to measure the understanding of fifth grade students of the primary school on the characteristics of wake and inter-wake relationship with basic competence to identify the properties of wake-up and investigate the nature of congruence and symmetry. Instrument of mathematics learning outcomes test consists of 30 question items by presenting 4 alternative answers. Measurement scale of the mathematics learning outcomes test is the score of 1 for the right answer and the score 0 for the wrong answer. Prior to the study all of these instruments have fulfilled the instrument feasibility test both the level of validity and reliability. Method of data analysis using two path or factorial 2 x 2 Anava. Before done data analysis, first do test requirement analysis that is test of normality and homogeneity.

Table 1. *Research design*

Variable X_2	Variable X_1	Learning Model	
		PBL (A_1)	Konventional (A_2)
	High (B_1)	A_1B_1	A_2B_1
Critical Thinking Skills	Low (B_2)	A_1B_2	A_2B_2

3. Findings andDiscussion

a. Test of Assumptions for ANOVA

The requirements of data analysis using parametric statistics are data obtained normal and homogeneous distributions. Normality test using kolmogorov-smirnov analysis and homogeneity test using Levene's test.

1) Normality test

Table 1 below presents result of normality test of data mathematics learning outcome through Problem Based Learning model and conventional model.

Table 2. *Result of Normality Test of Mathematics Learning Outcome Viewed from Learning Model*

	Learning Model	*Kolmogorov-Smirnov*			Decision	Conclusion
		Stat.	Df	Sig.		
Mathematics learning outcomes	PBL	0,068	153	0,082	Ho supported	Distributed normally
	Conventional	0,065	156	0,200	Ho supported	Distributed normally

Normality test with Kolmogorov-Smirnov analysis showed that the data of mathematics learning outcome through Problem Based Learning model and conventional model came from normal distributed population indicated by Sig. > 0.05.

The following table 2 shows the result of normality test of data mathematics learning outcome viewed High and Low Critical Thinking Skills

Table 3. *Result of Normality Test of Mathematics Learning Outcome Viewed from Critical Thinking Skills*

	Critical Thinking Skills	*Kolmogorov-Smirnov*			Decision	Conclusion
		Stat.	Df	Sig.		
Mathematics learning outcomes	Low	0,064	174	0,076	Ho supported	Distributed normally
	High	0,075	135	0,061	Ho supported	Distributed normally

Normality test by Kolmogorov-Smirnov analysis showed that the data of mathematics learning outcome viewed on high and low level of critical thinking skills came from normally distributed population, this is indicated by Sig. > 0.05.

2) Homogeneity Test

Table 3 below presents the homogeneity test of mathematicslearning outcomes viewed from learning model and critical thinking skills.

Table 4. *Results of Homogeneity Test of Mathematics Learning Outcomes Viewed from Learning Model and Critical Thinking Skills*

F	df1	df2	Sig.	Decision	Conclusion
1,662	3	305	0,175	Ho supported	Homogen

Results of homogeneity test show that the data of mathematics learning outcome viewed from learning model and critical thinking skills come from population having equal/homogeneous variance, this is indicated by Sig. > 0.05.

b. Two-Way Anova

Two-Way Anova Test aims to examine the effect of learning models and critical thinking skills and interaction effect of learning models and critical thinking skills on mathematics learning outcomes. Testing of Two-Way Anova using SPSS program summarized as the following table.

Table 5. *Result of Two-Way Anova*

Source	Sum of Squared	Df	Mean Squared	F	Sig.	Ho
Model	1454,743	1	1454,743	7,013	0,009	Supported
CTS	2270,971	1	2270,971	10,948	0,001	Supported
Model*CTS	22,940	1	22,940	0,111	0,740	Rejected

1) Differences in mathematics learning outcomes between students who following the learning of Problem Based Learning model with conventional model

Hypothesis 1 tested:

Ho: there is no difference in mathematics learning outcomes between students who following the learning of Problem Based Learning model with conventional model

H1: there are differences in mathematics learning outcomes between students who following the learning of Problem Based Learning model with conventional model

From the calculation of Anova (Table 4) obtained F-statistics 7.013 with Sig 0.009 <0.05 means Ho is rejected and H1 accepted. Means there is a significant difference in mathematics learning outcomes between students who following the learning of Problem Based Learning model with conventional model. Students who following the PBL model achieve higher learning outcomes than students who following conventional model.

2) Differences in mathematics learning outcomes between students who have high Critical Thinking Skills with Low Critical Thinking Skills

Hypothesis 2 tested:

Ho : there is no difference in mathematics learning outcomes between students who have high Critical Thinking Skills with Low Critical Thinking Skills.

H2: there is a difference in mathematics learning outcomes between students who have high Critical Thinking Skills with Low Critical Thinking Skills.

From the calculation of Anova (Table 4) obtained F-statistics 10,948 with Sig 0.001 <0.05 means Ho is rejected and H2 is accepted. This means that there are significant differences in mathematics learning outcomes between students who have high and low critical thinking skills. Students who have high critical thinking skills achieve higher learning outcomes than students with low critical thinking skills.

3) Interaction effect between learning model with critical thinking skills on mathematics

Hypothesis 3 tested:

Ho : there is no interaction effect between learning model with critical thinking skills on mathematics learning outcomes.

H3: there is interaction effect between learning model with critical thinking skills on mathematics learning outcomes

From the calculation of Anova (Table 4) obtained F-statistics 0,111 with Sig 0.740 > 0.05 means Ho accepted and H3 rejected. Means there is no interaction effect between learning model with critical thinking skills on mathematics learning outcomes. Thus, there is no difference in mathematics learning outcomes in both learning models in students with high or low critical thinking skills.

Based on the results of this study obtained the findings, that there are significant differences in mathematics learning outcomes between students who following learning of Problem Based Learning model with students who following conventional model. This is showed from Anova results with Sig. (0.009) <0.05. Based on the mean value of the mathematics learning outcomes of the two sample groups showed a significant difference that the mean value of the students following the Problem Based Learning model (73,68) was higher than the mean value of the students who following conventional model (67,92). Result of the observation showed the difference ofstudent'sbehavior, students activity in discussion, ability to analyze and solve the problem, liveliness to ask questions, activeness of doing the exercise questions for the experimental group students is higher than the control group students. Based on the findings of this study, PBL model emphasizes on learning by doing means students are invited to play an active role in learning activities because teachers give students the freedom to solve and learn ways or alternatives solving mathematical problems posed, so that students will be accustomed to solve critical problems critically and independent. The findings of this study support the results of previous studies conducted by Ajai et al. (2013) that there is a significant difference of post-test result of mathematics of students based on learning model, post test result of mathematics of students who follow PBL model learning is higher than conventional model. The findings of this study support the results of previous studies conducted by Fatade et al. (2013) that there is a significant difference of post test result of mathematics of junior high school students in Nigeria based on learning model, post test result of mathematics of students who follow PBL model learning is higher than control class students.

Based on the results of this study obtained the findings, that there are significant differences in the mathematics learning outcomes between students who have high critical thinking skills with students who have low critical thinking skills. This is evidenced from Anova results with Sig. (0.001) <0.05. Differences in mathematics learning outcomes of

students in terms of the high critical thinking skills and low critical thinking skills indicate different results on the tendency of students to perform cognitive processes that lead students to interpret, analyze, evaluate, inference, eksplanation, and self-regulation of a problem and mathematical problem solving. Thus, students cognitive ability to interpret, analyze, evaluate, conduct inference, explanation, and self-regulation is higher in students who have high critical thinking skills compared with students who have low critical thinking skills. The findings of this study support the results of previous research conducted by Fazriyah (2015) that there is a significant difference of science learning outcomes of fifth grade students in Primary School Depok viewed on critical thinking skills. Students who have high critical thinking skills, science learning outcomes of students given an integrative thematic learning model is higher than in the fragmented learning model. Students who have low critical thinking skills, higher learning outcomes of sciene use a fragmented model. The findings of this study support the results of previous research conducted by Erwin, Tellu and Kundera (2015) that there are significant differences in student learning outcomes in Biology lesson in SMAN 4 in Palu in terms of students critical thinking skills. Student learning outcomes that have high critical thinking skills is higher than students who have low critical thinking skills. The findings of this study support the results of previous research conducted by Nurwulandari, Muhardjito, and Mufti (2015) that there are significant differences in problem-solving abilities in terms of critical thinking skills in class X students in SMAN I Malang Year 2014/2015. Students with high critical thinking skills have higher problem-solving abilities than students with low critical thinking skills.

Based on the results of this study obtained the findings, that there is no interaction effect between learning models with critical thinking skills on mathematics learning outcomes. This is evidenced from Anova results with Sig. (0.740) > 0.05. Thus, simultaneously the factor of learning model and critical thinking skills do not give significant effect on students mathematics learning outcomes. In other words, learning model and critical thinking skills have influence partial on learning outcomes. To see and ascertain whether there is an interaction effect or not between learning models with critical thinking skills on mathematics learning outcome can be seen in the following graph:
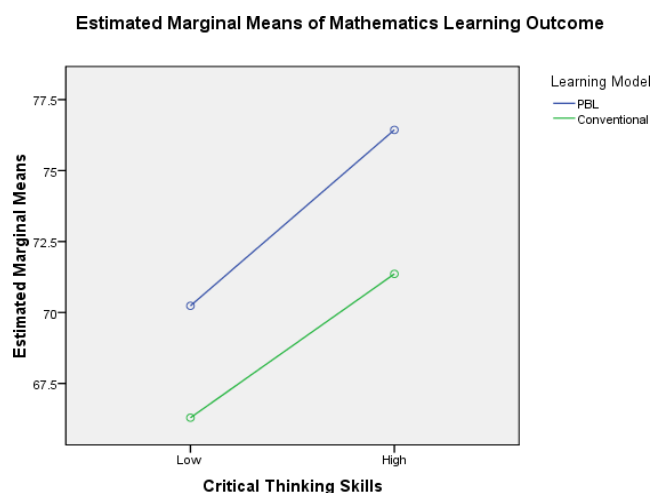

Figure 1. Graph of interaction effect between learning models with critical thinking skills on mathematics learning outcomes

According to Ghozali (2005: 59) Anova used to determine the main effect and interaction effects of independent variables (factors) on dependent variable. The main effect is the direct effect of independent variable on dependent variable, while the interaction effect is the effect of the two or more independent variables on dependent variable. To know the interaction effect other than can be seen from F statistic value and significance value, also can be seen from graph explaining the existence or absence of interaction between two factors. If two lines of the two factors intercept each other there is an interaction between the two factors. Conversely, if two lines of the two factors are not mutually intercepted there is no interaction between the two factors. From the results of data analysis with two-way anova generated graphs that describe applied of learning model to the two classes of Problem Based Learning model and conventional model with critical thinking skills that exist in students both high and low critical thinking skills. Based on the statement presented by Ghozali (2005: 59) can be obtained understanding, if there is an interaction between these two variables on mathematics learning outcomes, it will occur intersection between the graph of learning models with critical thinking skills. From the graph above, it can be seen that the two lines of both factors (learning model with critical thinking skills) do not cut each other, thus there can be no interaction effect between learning model and critical thinking skills on students matematics learning outcomes. In addition, each factor of learning model and critical thinking skills are not interdependent and influential but only have their own position on the learning outcomes. With no proven hypothesis three in this study can interpret that, this research model only use learning model and critical thinking skills as factor influencing mathematics learning outcomes. While outside the model of this study, there are still control variables that also influence the mathematics learning outcomes. Control variables that influence student learning outcomes can be sourced from internal factors such as students interest in learning, intelligence and external factors such as the use of learning media, learning conditions, and so forth. According to Gagne & Brings in Jamil (2012: 37) understanding of learning outcomes is the abilities that students have as a result of learning and can be observed through the performance of students. According to Anitah (2007: 27) the realization of behavioral changes of learning outcomes is an increase in the ability of students in accordance with the goals set. The success of achieving good learning outcomes in teaching and learning activities is influenced by two factors: the factors of within the students (intern) and factors from outside the student self (extern). The inner factors of the students that influence the learning outcomes include interest, talent, skill, effort, motivation, attention and health. Factors from outside students that influence the learning outcomes include physical and non physical environment. The physical environment in this case is learning tools, media, teachers and friends learn. The theory of learning methods proposed by Smith (in Jamil, 2012: 154) is an external factor that can affect student learning outcomes, the method of learning is the technical means taken by teachers to create teaching situations and can be developed for the implementation of learning with the aim to improve learning outcomes students. The theory proposed by Facione (2010: 5) explains that the aspect of critical thinking skills related to one's cognitive abilities is also an internal factor of students that can influence learning outcomes. The theory proposed by Chaffe (2012: 4) explains that critical thinking is the process of thinking to clarify one's understanding of something resulting in an intelligent decision. Based on these theories can be obtained understanding that the high low student learning outcomes can be influenced by many factors both internal factors and external factors. Based on the findings of this study, the absence of interaction influence between the two factors (learning model and critical thinking skills) on the result of learning mathematics can be caused because there are still many situational factors that can give effect (influence) on the improvement of learning outcomes both factors from within students as well as factors from outside the students. So based on these findings, the increase in

mathematical learning outcomes can be explained from the effects alone from both factors. The findings of this study support the results of previous research conducted by Erwin, Tellu and Kundera (2015) that there is no interaction between learning models and critical thinking skills of students on learning outcomes in Biology lessons in 4 senior high school of Palu. The findings of this study support the results of previous studies conducted by Widyatiningtyas et al. (2015) that there is no significant interaction between the learning models and the early ability of mathematics to the critical thinking skills of mathematics in high school students in Bandung. The findings of this study support the results of previous research conducted by Tijayanti and Marzuki (2014) that there is no significant interaction between the learning method and the type of intelligence on the development of students critical thinking skills in 1 senior high school of Suela Lombok Timur. The findings of this study support the results of previous research conducted by Ali (2012) that there is no interaction between learning approaches and the level of early knowledge of mathematics students to improve the critical thinking skills of mathematic of the junior high school students. The findings of this study support the results of previous research conducted by Risqi (2013) that there is no interaction between CTL and PBL methods with students creativity in learning mathematics on the achievement of seven grade students in junior high school of Muhammadiyah 5 Ngawi year 2013/2014.

4. Conclusions

Based on the results of research and discussion, can be concluded as follows (1) There are significant differences in mathematics learning outcomes between students who following Problem Based Learning model with conventional model of fifth grade students of the private primary school in Surakarta Region Year 2016/2017. Students who following Problem Based Learning model get higher learning outcomes than students who following conventional model (2) There is a significant difference in mathematics learning outcomes between students who have high and low critical thinking skills of fifth grade students of the private primary school in Surakarta Region Year 2016/2017. Students who have high critical thinking skills achieve higher learning outcomes than students with low critical thinking skills. (3) There is no interaction effect between learning model and critical thinking skills on mathematics learning outcomes of fifth grade students of the private primary school in Surakarta Region Year 2016/2017. There is no difference in mathematics learning outcomes in both learning models in students with high or low critical thinking skills.

Based on the conclusions from this study, teacher is suggested to improve regular meetings in the Teachers Working Group to discuss on the use of PBL model learning. Principal suggested supporting teachers in implementing learning innovation through workshop or in house training.

Further research is suggested to expand the research sample at public and private elementary school in Central Java area, so that research results can more be generalized. Furthermore, there is still a need for research to analyze the difference of mathematics learning outcomes by comparing between problem based learning with other model, so that the result will be more developed.

5. References

Ali, S. (2012). Improved Ability of Critical Thinking Mathematically Junior High School Students Through Contextual Teaching and Learning Approach. *Jurnal Edumatica*: *2*(1), 45-57.

Anitah, S. W. (2007). *Strategi pembelajaran di SD*. Jakarta: Universitas Terbuka.

Chaffe, J. (2012). *Thinking critically*. USA: Wadsworth Cengage Learning.

Erwin, M., Tellu, H.A.T., &Kundera, I.N.(2015). Comparison of Biology Learning Outcomes Taught Using Cooperative Integrated Reading and Composition (Circ) Model By Using Direct Learning Model Based on Critical Thinking Skill of Senior High Schoil Palu: *e-Jurnal Mitra Sains, 3*(1), 20-27.

Facione, P. A. (2015). *Critical thinking: what ıt ıs and why ıt counts*. 1-30. Retrieved from: http://www.insightassessment.com

Fatade, A. O., Mogari, D., & Arigbabu, A. A. (2013). Effect of problem based learning on senior secondary school students' achievements in further mathematics: *Internasional Journal Acta Didactica Napocensia, 6*(3), 27-49.

Fazriyah, N., Supriyati, Y., & Rahayu, W., (2017). The effect of integrated learning model and critical thinking skill of science learning outcomes: *Journal of Physics, 8*(2), 1-5.

Frank, Q. (2011). A science of learning approach to mathematics education: *Journal Notices of The American Mathematical Society, 58*(9), 1264-1275.

Ghozali, I. (2005). *Application of Multivariate Analysis With Spss Program*. Semarang: Diponegoro University Publishing Agency.

Jamil, S. (2012). *Learning Strategy, Theory and Application*. Yogyakarta: AR-Ruzz Media

John,T.A., Benjamin,I.I., & Emmanuel, I.O. (2013). Comparison of the learning effectiveness of problem based learning (PBL) and conventional method of teaching algebra. *Journal of Education and Practice, 4*(1), 131-136.

Johnson, E.B. (2009). *Contextual Teaching and Learning, Making Teaching and Learning Activities Fun and Meaningful*. Bandung : Mizan Media Utama.

Karim, N. (2015). Critical thinking ability in learning mathematics by using jucama model in junior high school. *JournalEducationMatemática*, *3*(1), 92-104.

Lian, I. T. A. (2010). Determining the readiness of staff for PBL training and development, *PBL Reflection-International PBL Symposium*, 9, 4-10.

Mazumder, Q. (2014). Student Motivation and Learning Strategies of Students from USA,, China and Bangladesh: *International Journal of Evaluation and Research in Education (IJERE)*, *3*(4), 205-210.

Moon, J. (2008).*Critical thinking an exploration of theory and practice*. USA: Routledge.

Murray-Harvey, R., Pourshafie.T., &Rayes, S. W. (2013). What teacher education students learn about collaboration from problem-based learning. *Journal of Problem Based Learning in Higher Education, 1*(1), 114-134.

Nurwulandari, N., Muhardjito, & Mufti, N. (2015). The influence of real-world applications ınquiry on the ability of problem solving of physical problems is the ability of thinking thinking critically in the X grade school. Proceedings of Scientific Meeting XXIX HFI Central Java and Yogyakarta, Yogyakarta April 25, 297-300.

Pecore, J.L. (2012). Beyond beliefs: teachers adapting problem-based learning to preexisting systems of practice. *Interdisciplinary Journal of Problem-Based Learning, 7*(2), 8-30.

Risqi, Y. (2013). Instructional learning using contextual teaching and learning (CTL) and problem-based learning (PBL) methodology is not soluble in the classroom and is used in the visualization of VIII SMP Muhammadiyah 5 Ngawi in the 2013/2014 genetics learning period. *Academic Journal and Maths*, *1*(3), 18-24.

Sanjaya, W. (2007). *Learning Strategies*. Jakarta: Kencana Prenada Media Group.

Talib, A. &Kailani, I. B., (2014). Problem based learning in cooperative situation (PBLCS) and its impact on development of personal intelligence. *International Journal of Evaluation and Research in Education (IJERE), 3*(4), 236-244.

Tijayanti,I., &Marzuki. (2014). The effectiveness of problem solving method in PKN learning for the development of the thinking ability of nationalism in SMA Negeri 1 Suela Lombok Timur: *SOCIA-JurnalIl IlmuSosial*, 2, 173-182.

Widyatiningtyas, R., Kusumah, Y. S., Sumarmo, U., &Sabandar, J. (2015). The impact of problem-based learning approach to senior high school students' mathematics critical thinking ability. *IndoMS-JME*, 2, 30-38.

IDENTIFYING UNIVERSITY STUDENTS' PERCEPTIONS OF 'ENGLISH' THROUGH METAPHORS

Dr. Cahit Erdem

Afyon Kocatepe University

cerdem@aku.edu.tr

Cahit ERDEM received BA degree in English Language Teaching department at Boğaziçi University, Turkey (2009). He holds a PhD in curriculum and instruction department at Anadolu University, Turkey. He is currently a lecturer at Afyon Kocatepe University, School of Foreign Languages, Turkey. His academic interests include media literacy, curriculum development, teacher education, educational technology, foreign language education and internationalization of higher education.

IDENTIFYING UNIVERSITY STUDENTS' PERCEPTIONS OF 'ENGLISH' THROUGH METAPHORS[1]

Cahit Erdem

cerdem@aku.edu.tr

Abstract

This study aims to identify university students' perceptions of English through the use of metaphors. This is a qualitative study with phenomenological design. The data were collected from 342 first grade university students at a state university through a semi-structured form, in which the participants were asked to complete the sentence "English is like… because…". The data were analyzed through content analysis method. Besides identifying positive, negative and neutral metaphors, the data were also analyzed in terms of participants' gender, faculty and high school type as well as forming categories of metaphors based on their rationales. Some of the main results of the study include; more than half of the students created negative metaphors, male and female participants' percentages of positive and negative metaphors are close to each other, the faculties with the higher frequency of positive metaphors are faculties of technology, fine arts and law while the faculties with the higher frequency of negative metaphors are faculties of medicine, engineering and veterinary medicine. 15 different categories of metaphors are formed. The results of the study are expected to guide practitioners by informing them on students' perceptions of English and help them adjust teaching-learning practices.

Keywords: English course, perceptions of English, metaphors

1. Introduction

Foreign language education is among the fields on which most emphasis is placed in higher education as well as elementary and secondary education. An intense foreign language education is provided to students at tertiary level particularly in compulsory or optional preparatory classes. In addition, compulsory courses of English (English-1 and English-2) are also offered to students at the first or second grade beside optional English classes such as academic writing or speaking. Despite this emphasis on teaching of foreign languages, particularly English, the quality of foreign language education in Turkey is criticized too often and reasons for the blamed failure are being researched (Atar, 2018; Cesur, 2008; Demir, 2012; Evcim, 2008; Memduhoğlu & Kozikoğlu, 2015).

There are numerous factors affecting the process of teaching and learning a foreign language such as characteristics of instructors, learning materials, educational requirements and so on; however, learners should be paid special attention as they are the main actors of this process. Learners' perceptions regarding foreign language affective characteristics are among significant factors in learning a foreign language and these factors include learners' levels of motivation, attitudes towards the language and their metaphors with respect to the target language (Gömleksiz, 2013). As learners' perceptions of the foreign language is one of the most important factors affecting the learning process, identification of these perceptions would be a guiding light for instruction practices. The use of metaphors is an efficient method in revealing learners' perceptions of foreign languages and their images in

---

[1] This study was presented in part at International Congress on Science and Education, 23-25 March, 2018, Afyonkarahisar, Turkey.

their minds (Limon, 2015). Metaphor is a reliable method to be used in measuring attitudes towards certain concepts (Coşkun, 2015) and commonly used in education research as it enables gathering rich data in small amounts and short time as well as making thorough interpretation on them (Koçyiğit, 2018).

Lakoff and Johnson (1982), the pioneers of the use of metaphors in research, assert that metaphors are used to understand and experience one kind of thing through another and add that they are mindsets that shape individuals' thoughts regarding the world and enable expressing ambiguous concepts in daily life in a more tangible manner. Therefore, metaphors have been used in the literature to enlighten tacit concepts. This trend is also common in the area of foreign language education. The main reason for this is the fact that identifying foreign language learners' perceptions regarding the target language is significant (Başaran-Uğur & Baysal, 2017). When identified, practitioners can make provisions for reversing negative perceptions and promoting positive perceptions.

Regarding tertiary level, a myriad of students receives foreign language education. Particularly, the compulsory foreign language courses for all departments in all universities address thousands of students. English is taught in these courses, which is also taught in elementary and secondary levels. Foreign language education is mostly in English in Turkey because it is accepted as the most prestigious language due to the policies of governments in Turkey (Sarıçoban & Sarıçoban, 2012). Due to predominance of English (Alptekin & Tatar, 2011), it is regarded as a synonym for foreign language education (Özmen, Cephe, & Kınık, 2016). Therefore, the quality of these compulsory courses needs to be researched and learners' perceptions of English is the focus of this study.

There are some research in the literature on foreign language education with respect to the use of metaphors; however, these studies are limited to the examination of perceptions of elementary and secondary school students towards 'English teacher' (Başaran-Uğur & Baysal, 2017; Ocak & Gürel, 2018), perceptions of high school students towards 'English' (Bekdaş, 2017; Limon, 2015; Ocak & Gürel, 2014), perceptions of university students towards 'English instructor' (Ahkemoğlu, 2011), 'foreign language' (Gömleksiz, 2013) and 'foreign language textbook' (Kesen, 2010). Perceptions of university students, particularly ones who take compulsory English classes, towards 'English' have not been examined, which would provide significant findings for the field and practitioners at tertiary level. Therefore, this study aims to identify first grade university students' perceptions of English through their use of metaphors. The research questions are as follows:

1. What is the distribution of metaphors in terms of positive, negative and neutral categories?

2. What is the distribution of positive, negative and neutral metaphors in terms of gender?

3. What is the distribution of positive, negative and neutral metaphors in terms of faculties?

4. What is the distribution of positive, negative and neutral metaphors in terms of high school type?

5. Under which categories do metaphors gather based on their rationales?

2. Method

2.1. Design

This qualitative study has been carried out in phenomenological design. This design includes examination of how people experience the world and the meanings people give to the things they experience (Newby, 2014). This study seeks to unearth the meanings first grade university students attach to 'English' through the use of metaphors.

2.2. Data Collection and Analysis

The data were obtained through a semi-structured form. The participants were asked to complete the sentence "English is like… because…" as well as demographic information part. The data were analyzed through content analysis method. The analysis process followed these steps:

- The appropriateness of the metaphors created by the participants was examined. 26 forms were excluded as they did not include a meaningful metaphor or had missing parts.

- The rationales of the metaphors (the part after 'because') were analyzed and they were grouped as positive, negative or neutral.

- Positive, negative and neutral metaphors were analyzed in terms of gender, faculty and high school type variables.

- Metaphors were grouped under new categories based again on their rationales.

- Specialist opinion was asked for the appropriateness of categories and the distribution of the metaphors under those categories, which made the results more valid.

2.3. Participants

The participants are 342 first grade students, 159 females and 183 males, attending to a state university in the Aegean region in Turkey. The participants were included in the research through convenience sampling method. The data were collected in the second and third weeks of fall semester of 2017-2018 academic year in order to reveal their perceptions before they took the course. 368 students in eight faculties took the form and 342 participants' forms were included in the analysis. Detailed information about the participants is provided in Table 1.

Table 1. *Participants' demographic information*

High school type	Faculty								Total
	Sci./Lit.	F. Arts	Law	Engineering	Health sciences	Tech.	Med.	Vet. Med.	
Open HS	0	1	0	0	0	1	1	0	3
Anatolian HS	26	35	16	32	41	60	17	24	251
Science HS	0	1	0	2	1	0	5	0	9
Religious Vocational HS	1	1	1	2	0	0	0	0	5
Vocational HS	7	6	0	2	0	8	0	0	23

Teacher Training HS	1	3	0	3	3	1	2	3	16
Medical vocational HS	0	2	0	0	7	0	0	0	9
International students	0	1	0	0	0	2	0	0	3
Basic HS	0	5	0	4	2	12	0	0	23
Total	35	55	17	45	54	84	25	27	342

3. Findings

3.1. The Distribution of Metaphors in Terms of Positive, Negative and Neutral Categories

The participants created 230 valid metaphors. 94 of them are positive and they are used 141 times. 146 metaphors are negative and they are used 182 times. 16 metaphors are neutral and used 19 times. The most frequently used metaphors include water (f=21), world (f=8), life (f=6), cat (f=6), obligation (f=6), torture (f=5), mathematics (f=5) and basic need (f=5). The most frequently used positive, negative and neutral metaphors are given in Table 2.

Table 2. *The most frequently used positive, negative and neutral metaphors*

Positive metaphors		Negative metaphors		Neutral metaphors	
Metaphor	*f*	*Metaphor*	*f*	*Metaphor*	*f*
Water	19	Cat	6	Flower	2
World	7	Torture	5	School	2
Basic need	5	Crossword puzzle	4	A new person	2
Telephone	4	Inessentiality	4		
Car	3	Obligation	4		
Fun	3	Sea	3		
Necessity	3	Prison	3		
Life	3	Life	3		
Money	3	Woman	3		
Total	141	Total	182	Total	19

The most recurring metaphors are shown in Table 2; however, there are also some metaphors used in different categories at the same time. The metaphors of riding a bicycle, insect, sea, world, literature, life, human, door, dream, meal, obligation are used both as positive and negative; the metaphor of song is used both as positive and neutral; the metaphor of doing sports is used both as negative and neutral; the metaphors of water, a little child and mathematics are used as positive, negative and neutral.

Of all the 342 metaphors, 41% is positive while 53% is negative. The percentage of neutral metaphors is 6. This finding reveals that negative perceptions are more dominant in the metaphors. Yet, the metaphors used also put forth that the participants' perceptions towards 'English' vary significantly. The most frequent metaphor is water. The participants refer that water is a basic and daily need and highlight the vitality of 'English'. The quotations of two participants are as follows.

"English is like water because it is indispensable in our lives just like water."

"English is like water because it is required throughout the life and we always encounter it."

The second most frequent metaphor is world. The students who have used this metaphor refer to the fact that English is a universal language and o common communication tool among people. Examples of this metaphors are as follows.

"English is like the world because it is used everywhere."

"English is like the world because we can talk and contact with people from all nations thanks to it."

3.2. The Distribution of Positive, Negative and Neutral Metaphors in Terms of Gender

The distribution of the metaphors in terms of gender is presented in Table 3.

Table 3. *The distribution of metaphors in terms of gender*

Gender	Positive		Negative		Neutral	
	f	%	F	%	f	%
Female	64	40	82	52	14	8
Male	77	42	100	55	5	3
Total	141	41 (all metaphors)	182	53 (all metaphors)	19	6 (all metaphors)

Of the 141 positive metaphors, 64 were created by females and 77 were created by males; of the 182 negative metaphors, 82 were created by females and 100 were created by males; of the 19 neutral metaphors, 14 were created by females and 5 were created by males. As the number of males and females isn't equal, the percentages were calculated. Of the metaphors created by females, 40% are positive, 52% are negative and 8% are neutral while of the metaphors created by males, 42% are positive, 55% are negative and 3% are neutral. These findings reveal that the percentages of positive and negative metaphors are quite close to each other for males and females, which suggests that the perceptions of students do not differ by gender to a great extent. It should also be noted that more than half of the both males and females have negative perceptions regarding 'English'.

3.3. The Distribution of Positive, Negative and Neutral Metaphors in Terms of Faculty

The distribution of the metaphors in terms of faculties students attend to is presented in Table 4.

Table 4. *The distribution of metaphors in terms of faculties*

Faculty	Positive		Negative		Neutral	
	f	%	f	%	f	%
Science and literature	12	34	19	54	4	12
Fine arts	25	45	27	49	3	6
Law	7	41	9	53	1	6
Engineering	13	29	30	67	2	4
Health sciences	19	35	31	58	4	7
Technology	52	62	31	37	1	1
Medicine	5	20	18	72	2	8

Veterinary medicine	8	30	17	63	2	7

An analysis of the distribution of metaphors by the faculties reveals that the students with the highest number of positive metaphors belong to technology faculty (62%). Students at the faculty of fine arts (45%) and faculty of law (41%) follow it. The students with the highest number of negative metaphors belong to the faculty of medicine (72%). Students at the faculty of engineering (67%), faculty of veterinary medicine (63%) and school of health sciences (58%) follow it. Among neutral metaphors, the highest percentage belongs to faculty of science and literature (12%) and the lowest percentage belongs to faculty of technology (1%).

The distribution of metaphors by the faculties suggests that, except for faculty of technology, the students studying at departments of social sciences tend to create more positive metaphors while students at departments of hard (quantitative) sciences tend to create more negative metaphors, which unearths the need to analyze high school types.

3.4. The Distribution of Positive, Negative and Neutral Metaphors in Terms of High School Type

The distribution of the metaphors in terms of high school types that students graduated from is presented in Table 5.

Table 5. *The distribution of metaphors in terms of high school types*

High school type	Positive		Negative		Neutral	
	f	%	f	%	f	%
Open high school	1	33	2	67	0	0
Anatolian high school	113	45	124	49	14	6
Science high school	3	33	5	56	1	11
Religious vocational high school	0	0	3	60	2	40
Vocational high school	9	39	14	61	0	0
Teacher training high school	5	31	10	63	1	6
Medical vocational high school	1	11	8	89	0	1
Basic high school	7	30	15	66	1	4
International students	3	100	0	0	0	0

As the number of participants is not equal among groups, the percentages are considered. The students with the highest number of positive metaphors are graduates of Anatolian high schools (45%). Graduates of vocational high schools follow it (39%). None of the graduates of religious vocational high school students used positive metaphors and all of the international students used positive metaphors.

The highest percentage of negative metaphors belongs to graduates of medical vocational high school (89%). That nearly all of those students created negative metaphors puts forth their negative perceptions towards 'English'. Percentage of negative metaphors is also quite high with graduates of open high school, basic high school, teacher training high school and religious vocational high school respectively.

Neutral metaphors have low level percentages, except for graduates of religious high school students (40%), which may have stemmed from their low number.

3.5. Categories of Metaphors Based on Their Rationales

Based on the rationales students provided in their sentences after the word 'because' in the form, 15 different categories were created after an elaborate grouping process. The categories and sample metaphors are provided in Table 6.

Table 6. *Categories of metaphors based on their rationales*

Categories and related metaphors	f	%
Difficulty	101	29
Soup, nightmare, black hole, chaos, summer Quran course, deadlock, a high mountain, dead well, night terror, cat, sudoku, torture, swing carousel, wall, prison		
Necessity	54	16
breath, water, basic need, oxygen, telephone, puzzle piece, life, meal, foot, friend, right hand, mother, handkerchief, drug, pasta in a student's home, book, insect		
Requiring attention/effort	32	9
pet, flower, a little child, a child crying at night, motor vehicle, cat, patient, plant, a delicate lover, woman, grass, doing sports, nature, frozen water		
Communication tool/ cultural contribution	22	6
car, freedom, a full bus ticket, human, world, eye glasses, discovering a new place, bridge, passport, a new life, ship, telephone, pen, train, plane, culture		
Self-improvement	18	5
bread, life, expressing oneself, discovery, a second brain, telephone, walking, horse, knowledge, sea, door, family, tree, computer, riding a bicycle, a new friend		
Dislike	18	5
pumpkin, lumber, lesson, Çorum, Afyon, Aleyna Tilki, baklava, grapefruit, our shadows, torture, insect, Mondays, cigarette, depression, boredom		
A tool with gains	17	5
leadership, golden bracelet, car, step, highway, money, key, gun		
Obligation	15	4
A jumper given as a present, our shadow, prison, boomerang, identity card, organ, torture, human		
Easy to forget	13	4
Fish memory, butterfly, cat, dream, air, pencil, writing on the water, flying off ink, flower		
Fun	11	3
waterfall, fun, raspberry cake, hidden garden, funfair, color, train		
Loving	9	3
literature, dream, a little child, happiness, therapy, tv series		
Disappointment	9	3
love, candy apple, life, doing sports, fruit gum, hot pepper, wind		
Universality	8	3
partnership, water, globe, world, translator		
Futility	8	3
shower gel, communism, nonsense, stone, trouble, an empty box		

Dislike but necessity	7	2
garlic, school, water of the pool, roquette, manure, chicken soup, vegetable dishes		

While forming the categories, it was tried to form them as comprehensive as possible. For instance, the categories of 'repeating' and 'obstacle' were also formed in the first analysis; however, they were included in the category of 'difficulty'. Similarly, the category of 'boring' was added to 'dislike' category. Yet, the categories of 'fun, disappointment, easy to forget, requiring attention/effort' were not included in other related categories because they were particularly highlighted in the rationales though they overlapped with other categories.

The distribution of categories reveals that the majority of the metaphors created by the students fall into the category of 'difficulty' (f=101). This puts forth that a greater part of the students perceives 'English' as a difficult language or learning process. Regarding difficulty of 'English', some of the quotations of the students are as follows.

"English is like a nightmare because both are scary".

"English is like getting lost because no matter for how long I have been learning English, I am still getting lost in the sentences".

"English is like a high mountain because you can never reach to top level".

"English is like cooked rice in the plate because you start eating but can never finish it".

'Difficulty' is followed by the category of 'necessity' regarding frequency of use (f=54). This implies that although students perceive 'English' as a difficult language, they are aware of its significance and necessity. Sample quotations in this category are as follows:

"English is like oxygen because it is a language that we should learn and we need just like oxygen".

"English is like a basic need because it is needed in nearly whole world".

"English is like a telephone because it is a part of our lives like a telephone".

On the other hand, seven students stated they did not like 'English' but they knew that it was necessary. Samples are provided below. Most of the metaphors in this regard are about food the participants do not like.

"English is like a chicken soup because it is nutritious but does not taste good".

"English is like roquette because it is disgusting but has many benefits".

"English is like vegetable dishes because we need to take it yet it turns my stomach".

Eight students see English or learning English as a futile activity, as stated in the following quotations.

"English is like a stone because it has never had a benefit for me like a stone on the street".

"English is like a nonsense because there is no need for English as we have Turkish. Will I work abroad? I will work in Turkey".

The categories of 'requiring attention/effort (f=32) and 'easy to forget' (f=13) are related to the category of 'difficulty'. A considerable number of students stated English was easy to forget and it requires hard work for this reason. Sample quotations in the category of 'easy to forget' include following.

"English is like a dream because I take it every year and I forget half of what I have learnt at the end of that semester".

"English is like a butterfly because we forget what we learn in a day like a butterfly".

Sample quotations in the category of 'requiring attention/effort' are as follows.

"English is like a little child because it needs continuous attention".

"English is like a motor vehicle because just like how a motor vehicle cannot be used without oil and there occurs a problem if we do not carry out its maintenance, we cannot move on with English if we do not memorize vocabulary and do not speak it even if we can".

"English is like a cat because if you do not care about it shows ingratitude and runs away".

'English' is also considered as a 'a tool with gains', 'communication tool/ cultural contribution' and a necessity for 'self-improvement' by students. There are also students who stated they did not like 'English' without providing a reason.

The categories can also be labeled as positive, negative and neutral. The positive category includes the categories of 'necessity, communication tool/ cultural contribution, self-improvement, a tool with gains, fun, loving and universality' while the negative category includes 'difficulty, requiring attention/effort, dislike, obligation, east to forget, disappointment, futility'. 'Dislike but necessity' can be considered as neutral.

4. Discussion and Conclusion

Identifying first grade university students' perceptions towards 'English' is a significant issue as the compulsory English classes at tertiary level address thousands of students. Despite this fact, the students taking these classes are mostly ignored in the literature. To this end, 342 first grade university students' metaphors of 'English' were analyzed. The most frequent metaphors include 'water, world, life, cat, obligation, torture, mathematics and basic need'. The analysis reveals that 240 valid metaphors were created by 342 students, 53% of which are negative (146 out of 240). The percentages of positive and neutral metaphors are 41% and 6% respectively. This finding suggests that more than half of the students have negative perceptions regarding 'English' and it is in line with Bekdaş (2017) and Limon (2015) while it differs from Gömleksiz (2013). With high school students, Bekdaş (2017) and Limon (2015) analyzed metaphors of 'English' and they found out that 52% and 57% of the metaphors were negative. On the other hand, Gömleksiz (2013) reported surprising results as to the metaphors of prospective teachers regarding the concept of 'foreign language'. He found out that 80% of the participants created positive metaphors. Except for this study, the literature supports the finding in the current research and suggests that the negative attitudes towards English need to be taken into account for an effective foreign language teaching and learning experience.

It was found out in the study that males and females' percentages of positive and negative metaphors are quite close to each other. 40% of females and 42% of males created positive metaphors while these percentages are 52 and 55 for negative metaphors. This result is also supported by Bekdaş's (2017) study and suggests that there is not a high level of difference between males and females in this respect.

With respect to faculties, the highest percentages of positive metaphors were created by students at technology, fine arts and law faculties and the highest percentages of negative metaphors belong to faculties of medicine, engineering, veterinary medicine and school of health sciences. This finding poses that the students studying at departments of social sciences tend to create more positive metaphors while students at departments of hard

(quantitative) sciences tend to create more negative metaphors, except for faculty of technology where the instructor talked about the significance of English in the first week and this may have contributed to the high percentage in this faculty. Faculties such as medicine and engineering accept students with quantitative scores from university entrance exam and it can be argued that students in these departments tend to ignore subjects like English, as supported by the findings. These findings are also supported by the analysis of high school types that students had graduated from. The highest percentage of negative metaphors belongs to graduates of medical vocational high schools, which accounts for the high percentage of negative metaphors in school of health sciences. Another surprising finding in this issue is that graduates of vocational high schools created more positive metaphors than graduates of science high schools and teacher training high schools. This unearths that foreign language education at tertiary level cannot be considered as independent of education at former levels. Besides, students aiming high academic achievement, particularly at quantitative departments, do not place emphasis on 'English' and students who have created rationales under the theme of 'futility' are all students at quantitative departments. It should also be noted that no metaphors study analyzing students' faculties and high school types could be found in the literature.

Based on the rationales provided by the students, categories were formed. The most frequent category is 'difficulty'. A great number of students perceive 'English' as a difficult language, which interferes with learning process. Despite seeing it as a difficult language, a considerable part of the students accepts English as a necessity, which is good for language learning. 'Water', the most frequent metaphor, includes rationales in the 'necessity' category. Related to the 'difficulty' category are the categories of 'requiring attention/effort' and 'easy to forget'. The repetition of the subjects every year, forgetting what has been learned in a short time, the lack of speaking skills and focusing on the teaching of grammar are present in the rationales. Analyzing all categories together, it can be deduced that students are aware of the necessity in learning English and they even see it as a communication/ cultural contribution tool with gains; however, the majority of students regard it as a difficult language. This percentage gets bigger when the categories of 'requiring attention/effort, easy to forget and disappointment'. This perception of difficulty also seems in parallel with public opinion (Limon, 2015), which should be taken into account by the practitioners.

Though any study on the university students' metaphors of 'English' could not be found, the categories in similar studies seem to overlap with the categories created in this study. High school students' metaphors of 'English' (Limon, 2015) and prospective teachers' metaphors of 'foreign language' (Gömleksiz, 2013) were analyzed by the researchers and their categories are compared. Firstly, it should be noted that, as explained in detail in findings section, the categories in this research were formed comprehensively; however, certain categories which are also related to other categories are kept as separate if it is highlighted by the students in the rationale sections. Therefore, this study has a more elaborate set of categories. The categories of 'a difficult and complicated language', 'necessity', 'a hard and dull process', 'result of accumulation', 'a functional tool', 'a funny activity', 'a valuable language' and 'a futile and ineffective process' in Limon's (2015) study corresponds to the categories of 'difficulty', 'necessity', 'requiring attention/effort', 'communication tool/cultural contribution', 'a tool with gains', 'fun', 'loving' and 'futility' in this study respectively. Similarly, the categories of 'difficulty', 'necessity', 'power', 'future', 'pleasure and travel', 'despair' and 'colonialism' in Gömleksiz's (2015) study corresponds to the categories of 'difficulty', 'necessity', 'self-improvement, a tool with gains', 'fun', 'disappointment' and 'futility' in this study respectively. The categories of 'difficulty',

'necessity', 'a tool with gains', 'fun' and 'futility' are common in all three studies. This reveals a similar pattern in students' perceptions of English.

5. Recommendations

This study puts forth university students' perceptions of English through metaphors, which has yielded significant results, particularly for practitioners. They can take various precautions taking the categories into account. For instance, the study identified that the majority of students perceive English as a difficult language or a number of them do not like the language itself. Therefore, the instructors can try to break down these prejudices with appropriate activities. In the context of this study, the instructors in the faculties of medicine, engineering, veterinary medicine and school of health sciences need to try to eliminate negative perceptions towards the language. In addition, it should be noted that the use of metaphors is an easy to implement and analyze method to be used by teachers or instructors of a foreign language. Using this method at the beginning of a semester would help them guide the learning process efficiently. More studies are needed in this respect to better understand university students' perceptions of English or any other foreign language with respect to different concepts and variables to better manage foreign language education at tertiary level. Curriculum development and instructional design studies can also benefit from the results of these studies particularly in selecting and organizing contents of the curricula by aligning the contents with students' perceptions.

References

Ahkemoğlu, H. (2011). *A study on metaphorical perceptions of EFL learners regarding foreign language teacher* (Unpublished master dissertation). Çukurova University, Institute of Social Sciences, Adana, Turkey.

Alptekin, C. & Tatar, S. (2011). Research on foreign language teaching and learning in Turkey (2005-2009). *Language Teaching, 44*(3), 328-353.

Atar, C. (2018). Incidental learning or intentional learning: A compromising and complementary account. In C. Atar (Ed.), *Cutting-edge topics and approaches in education and applied linguistics*. Newcastle: Cambridge Scholars Publishing.

Başaran-Uğur, A. R., & Baysal, S. (2017). The metaphorical perceptions of secondary school students regarding the concept of English teacher. *The Journal of International Social Research, 10*(52), 730-736.

Bekdaş, M. (2017). The perceptions of high school students for the concept of English through metaphors: Tokat sample. *Gazi Journal of Education Sciences, 3*(2), 35-41.

Cesur, M. O. (2008). *A model explaining and predicting the relationship between university prep class students' language learning strategies, learning styles and academic success in foreign language* (Unpublished doctoral dissertation). Yıldız Technical University, Institute of Educational Sciences, İstanbul, Turkey.

Coşkun, A. (2015). Parents and young learners' metaphorical perceptions about learning English. *Journal of Education and Training Studies, 3*(5), 231-240.

Demir, Y. (2012). *A comparative analysis on sofl students' perceptions regarding nest and non-nest effects on motivation and attitudes* (Unpublished master dissertation). Selçuk University, Institute of Educational Sciences, Konya, Turkey.

Evcim, H. (2008). *Increasing vocabulary retention of university prep class students through language learning strategies* (Unpublished master dissertation). Abant İzzet Baysal University, Institute of Social Sciences, Bolu, Turkey.

Gömleksiz, M. N. (2013). Metaphorical perceptions of prospective teachers regarding foreign language. *Turkish Studies, 8*(8), 649-664.

Kesen, A. (2010). Turkish EFL learners' metaphors with respect to English language coursebooks. *Novitas Royal Research on Youth and Language, 4*(1), 108-118.

Koçyiğit, M. (2018). Eğitim bilimsel temelleri [Scientific foundations of education]. In A. Arı (Ed.), *Eğitim bilimine giriş [Introduction to educational sciences]* (p. 209-220). Konya: Eğitim Publishing.

Lakoff, G., & Johnson, M. (1982). *Metaphors we live by.* Chicago, IL: University of Chicago Press.

Limon, İ. (2015). 9th and 10th graders metaphorical perceptions about English as a foreign language. *International Journal of Social Sciences and Education Research, 1*(2), 367-379.

Memduhoğlu, H. B., & Kozikoğlu, H. (2015). The attitudes of university students concerning foreign language courses. *Dicle University Journal of Ziya Gökalp Education Faculty, 24*, 184-202.

Newby, P. (2014). *Research methods for education.* Abingdon: Routledge.

Ocak, İ., & Gürel, E. (2014). Lise öğrencilerinin İngilizce dersine yönelik metaforik algıları [High school students' metaphorical perceptions regarding English class]. *23. Educational Sciences Congress,* 4-6 September, Kocaeli, Turkey.

Ocak, İ., & Gürel, E. (2018). İlköğretim 4. 5. 6. 7. 8. sınıf öğrencilerinin İngilizce öğretmeni yönelik metaforik algıları [Metaphorical perceptions of elementary students at 4. 5. 6. 7. 8. Grades towards English teacher]. *International Congress on Science and Education 2018*, 23-25 March, Afyonkarahisar, Turkey.

Özmen, K. S., Cephe, P. T., & Kınık, B. (2016). Trends in doctoral research on English language teaching in Turkey. *Educational Sciences: Theory and Practice, 16*(5), 1737-1759.

Sarıçoban, G., & Sarıçoban, A. (2012). Atatürk and the history of foreign language education. *The Journal of Language and Linguistic Studies, 8*(1), 24-49.

FOREIGN LANGUAGE TEACHING ANXIETY AMONG PRE-SERVICE TEACHERS DURING TEACHING PRACTICUM

Dilek Tüfekçi Can

Balıkesir University

tufekci@balikesir.edu.tr

Dilek Tüfekçi Can is an assistant professor at the ELT Department of Balıkesir University. She has many published research articles in international journals. Her research interests include teaching English to young language learners, pre-service teacher education, linguistics, discourse and literature.

FOREIGN LANGUAGE TEACHING ANXIETY AMONG PRE-SERVICE TEACHERS DURING TEACHING PRACTICUM

Dilek Tüfekçi Can

tufekci@balikesir.edu.tr

Abstract

Research on anxiety in English as a Foreign Language (EFL) commonly focuses on language learning and its related matrices whereas foreign language teaching anxiety (FLTA) still remains an area of research to be addressed and explored thoroughly. Accordingly, this study was carried out to fill the gap in the literature since when the literature was reviewed, it was found that the few definitions on FLTA were inadequate in covering many other aspects of the concept itself. Thus, the aim of this study was firstly to define a newly emerging concept named FLTA. Secondly, the study aimed to explore pre-service EFL teachers' anxiety provoking factors while they were practising teaching English to young learners at real classroom settings during teaching practicum. The main aim of this study was to make a novel classification on anxiety provoking factors, namely the sources of FLTA. The study was carried out with 25 pre-service EFL teachers studying in the ELT department at Balikesir University. A background questionnaire, interviews, reflections and essay papers were used to collect qualitative data. The results indicated that the pre-service EFL teachers experienced anxiety in teaching English to young learners at real classroom settings particularly related to cognitive, affective and socio-cultural factors.

Keywords: English as a foreign language, pre-service teachers, teaching anxiety, teaching practicum

1. Introduction

Research on anxiety has gained a growing prominence in all spheres and levels of education, particularly in the departments of English Language Teaching in the past decades. The growing concern on anxiety in the field of language learning/teaching mainly stems from the emergence of the new studies on the crucial impacts of affective factors on foreign language learning/teaching processes. Since anxiety is commonly considered as an obstacle for learners which impedes their ability to perform the language productively in foreign language learning, numerous studies have been conducted to examine the role of anxiety on foreign language learning and, its causes and effects on learning. Whereas the research on foreign language anxiety (FLA) has significantly increased in the last decades, foreign language teaching anxiety (FLTA) has remained a rather untouched area of research. As an indication to this, Tum (2012) stated that the research carried out on learning anxiety outweighed the research on FLTA in number and also revealed that the research on FLTA remained rather limited to draw satisfactory conclusions on the related subject.

When the related literature was reviewed, some definitions on the anxiety in general are as noted in the following part. Anxiety is commonly defined as 'something felt' by Freud (1924). In a more specific definition, it is a kind of unpleasant emotional state or condition that includes feelings of uneasiness, worry and a psychological excitement, and is equated with objective anxiety (Freud, 1936), an anxiety which is considered as an emotional reaction with its proportional intensity to a real danger in the external world. As an affective state,

anxiety is classified into three types as *trait*, *state* and *situation-specific anxiety*. *Trait anxiety*, a kind of constant personality trait, is heavily dealt with an individual's inborn qualities, namely an aspect of personality (Scovel, 1978). *State anxiety*, as a transient anxiety, is regarded as an emotional reaction or response to a particular situation (Spielberg et al., 1983) such as an important test. *Situation-specific anxiety* is associated with explicit situations and events (Horwitz et al., 1986) such as public speaking, examinations, or class participation (Ellis, 1994).

Language Anxiety is defined as 'a construct of anxiety which is not general but instead is specific to the language acquisition context is related to second language achievement' (Gardner, 1985, p. 34). Young also states that 'language anxiety is a complex psychological construct requiring investigation from a variety of perspectives and approaches' (1992, p. 157). However, *foreign language anxiety* (FLA) is described as 'a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning experience' (Horwitz et al., 1986, p. 128). Yet again, some researchers (Baltacı, 2017; Özturk, 2016) stated that there were not any identifiable definitions of FLTA in the literature. Moreover, Numrich (1996) generally defined FLTA as a part of problems experienced by language teachers. Furthermore, some other attempts were also made to define the FLTA as "an emotional and affective state that a teacher feels tension due to personal, perceptional, motivational and technical concerns before, during and after teaching activities" (Aydin, 2016, p. 639). However, all these definitions, which can be regarded as among the first attempts made in the literature, become rather superficial as they restrict the term into somewhat emotional and affective states, and as they confine the concept either only to the process of teaching activities or teachers themselves. Moreover, these definitions also lack cognitive and socio-cultural states of the teachers. Thus, a more comprehensive definition is suggested by the researcher of this paper.

As seen, the definitions made in the literature above are commonly about FLA and FLTA has not been fairly defined in the current literature. Thus, the following definition can be regarded as another attempt to fill the gap in the literature as the emergence of this definition owes itself to the oral reassurance of the distinguished researcher Prof. Dr. Rod Ellis himself in the 4th & 5th International Conferences on New Trends in English Language Teaching and Testing which was held in Istanbul, on March 30, 2018. *Foreign language teaching anxiety* (FLTA), a kind of precisely *situation-specific anxiety* as Ellis (1994) labelled, can be defined as a psychological discomfort or apprehension which is unique for pre-service teachers, language teachers, instructors and practitioners who are responsible for teaching learners how to learn a foreign language effectively and professionally at a real classroom setting by using all skills and subskills required for teaching the target language. In other words, FLTA, which stems from cognitive, affective and socio-cultural factors, can also be defined as the feeling, believing and thinking of being inadequate in teaching and practising the foreign language as well as of being incompetent in performing the language acceptably with the real students at a real teaching context.

In sum, the present study was carried out to reveal some significant factors on FLTA. Firstly, anxiety in foreign language context was mainly researched in learning context, which is mostly a simulation as is the case in the study of Aydin (2016) rather than a real class setting in a teaching context. Thus, this research was carried out in a real classroom setting as a teaching context in order to reveal the role of contextualization on FLTA. Secondly, this paper was an attempt to (re)define the term FLTA in literature and to attract the attention of the researchers on making a discrimination between FLA and FLTA. Thirdly, this study also attempts to make a holistic classification rather than an atomistic one as it classifies anxiety provoking factors into subtitles such as cognitive, affective and socio-cultural factors. Thus,

this study will bring a new perspective on anxiety-provoking factors in terms of its taxonomy. Moreover, since the researcher of this paper who has been working at Foreign Languages Teaching Department, particularly at English Language Teaching Department for more than 15 years observed the pre-service EFL teachers' anxiety on teaching rather than learning when she was lecturing on both Teaching Experience and Teaching Practicum. With these concerns in mind, this study aimed to investigate the anxiety-provoking factors among pre-service EFL teachers while teaching English to young language learners at a real classroom setting as a part of teaching practicum.

2. Literature Review

In the literature review, it must be noted that FLA, which is generally regarded as a universal phenomenon that inhibits students' achievement, both has a debilitating and facilitating effect on foreign language learning. For instance, whereas a number of research on FLA revealed the debilitating effect of anxiety on learning and achievement (Horwitz & Cope, 1986; MacIntyre & Gardner, 1989; MacIntrye & Gardner, 1994; Oxford, 1999; Trylong, 1987; Young, 1991), some of the others indicated the facilitating effect of anxiety in learning a foreign language (Alpert & Haber, 1960; Chastain, 1975; Scovel, 1978). Some studies (Fletcher, 1997; Hancock, 2001) also indicated that anxiety generates an adverse effect on performing foreign language.

It is thought that the nonnative pre-service EFL teachers can be more prone to FLTA. Tum (2015) hypothesized that FLTA may stem from three periods. First, the pre-service EFL teachers may experience FLA themselves as language learners in the foreign language classrooms. Second, as pre-service teachers, they become more aware of the challenges and responsibilities awaiting them in their future profession as they came closer to their training course. Moreover, entering the foreign language classroom for the first time as a teacher may seem overwhelming for relatively inexperienced teachers.

In the literature on FLTA, Merç (2004) found that anxiety was one of the most frequently reported problems of pre-service EFL teachers completing their teaching practicum at Anadolu University Faculty of Education. Moreover, Kim and Kim (2004) also conducted a study focusing on the most anxiety-provoking sources for pre-service teachers. The results revealed that the pre-service teachers were anxious at times when teaching English through English; teaching English to the students who had lived in English speaking countries; teaching unmotivated students; teaching speaking; teaching English listening; being asked unexpected questions; being unable to control class and lastly, being observed in classes. The study also outlined the sources of FLTA as limited English proficiency, lack of confidence, lack of knowledge about linguistics, insufficient class preparation, being compared to native teachers, fear of negative evaluation, and lack of teaching experience. Furthermore, İpek (2007) revealed six categories of sources of anxiety: making mistakes, teaching a particular language area, using the native language, teaching students at particular language levels, fear of failure, and being compared to fellow teachers. Merç (2011) also conducted another study on the sources of anxiety experienced by pre-service EFL teachers throughout teaching practicum and categorized the sources of FLTA as students and class profiles; classroom management; teaching procedures; being observed; mentors and miscellaneous.

Horwitz (1996) claimed that foreign language teaching anxiety can deter teachers from using target language productively, interacting with the students and being a positive role model as a language learner. In a similar vein, Öztürk (2016) revealed that teachers possess a mediocre level of teaching anxiety while teaching a specific skill in which they feel themselves incompetent. The factors causing teaching anxiety were classified into three titles as knowledge of target language items, students' manners, and knowledge of target culture.

Like Öztürk's (2016) study, the study of Fish and Fraser (2001) showed that the activities related to teaching were the sources of the highest levels of teaching anxiety. Similarly, the study of İpek (2016) revealed that non-native EFL teachers experienced anxiety specific to teaching the target language when they made mistakes; felt inadequate in teaching grammar; used the native language; taught students at particular language proficiency levels; had a fear of failure.

Among the studies conducted on the sources of FLTA, the study of Na (2007) indicated the variables of anxiety provoking factors as communication apprehension, fear of negative evaluation, test anxiety, anxiety of English classes and English classroom anxiety. Whereas the study of Gardner and Leak (1994) specified the sources of teaching anxiety in relation to academic rank and teaching experience, Yoon (2012) identified the sources as the use of target language, lack of self-confidence and lack of preparation. Paradoxically, Aslrasouli and Vahid (2014) examined anxiety among EFL teachers in the Iranian context and revealed different sources of anxiety among female and male teachers such as interpersonal relations, employment structure, language proficiency and knowledge, facilities and resources, and other factors. In another research (Thomson & Lee, 2013), the effect of language anxiety on achievement was examined and an intriguing new factor, namely the fear of ambiguity in English emerged as a new factor in terms of language anxiety. In a study conducted in Turkish context (Tum, 2015), the results indicated that anxious pre-service teachers experienced significant levels of language anxiety to a degree that may cause them to avoid using the target language and language-intensive teaching practices in their classrooms.

FLTA also stems from some of the other factors such as the identity-based anxiety (Stroud & Wee, 2006), the causes of which are fear, shyness and discomfort (Anandari, 2015); the students' indifference in taking risk using/speaking English in class, the students' fear of being negatively evaluated and being apprehensive about public speaking and tests; the students' unwillingness to communicate (Liu & Jackson, 2008); the teachers' inadequate attention to anxiety (Trang, Baldauf, & Moni, 2013); the fear of ambiguity in English (Thomson & Lee, 2013) and the students' country of origin (Lim, 2009).

In sum, the classifications made on anxiety provoking factors are rather specific classifications than a general one. In contrast to atomistic approach, this research on anxiety provoking factors or the sources of teaching anxiety mainly bases its premises on holistic approach and specifically attempts to make a general classification on FLTA such as cognitive, affective and socio-cultural factors.

3. Methodology

3.1. Research Design

The present study employed qualitative research design, which included data collection procedures that generally resulted in open-ended and non-numerical data (Dörnyei, 2007). It used strategies of inquiry such as narratives, phenomenology, and case studies (Creswell, 2003). As a case study, this research used a descriptive method as it focused on anxiety provoking factors experienced by pre-service EFL teachers. The research aimed at answering the following research questions:

1. What are the pre-service teachers' opinions on teaching practicum in general?

2. Do the pre-service EFL teachers feel teaching anxiety during teaching practicum?

3. What are the anxiety-provoking factors during teaching practicum?

3.2. Participants

The sample group in the study consisted of 25 pre-service EFL teachers, 7 males and 18 females. All participants were fourth year students at the ELT department. The mean age was 23.6, in the range of 21 to 41. Of the participants, 28% (7) were males whereas 72% (18) were females. They were enrolled in the Teaching Practicum course as a requirement for their graduation in 2016-2017. The research lasted 14 weeks in English Language Teaching Department at Balikesir University in Turkey. During the research, the pre-service EFL teachers were observed by the practice teachers particularly when they delivered their lessons. They were also observed and assessed by the university supervisors at least two times during their performances.

3.3. Data Collection Instruments

The data collection instruments of this qualitative study consisted of a background questionnaire, interviews, reflection and essay papers. The background questionnaire was composed of open-ended questions about the pre-service EFL teachers' opinions on anxiety-provoking factors. The open-ended questions, which were an important part of semi-structured interviews in order to obtain responses from the participants who were using their own words, rather than the pre-defined ones by the researcher or past findings (Yin, 2010), were instantaneously raised in connection with the study. Thus, the questionnaire centered on examining pre-service EFL teachers' opinions developed along the interaction with the process of teaching. Secondly, in the interviews, as a qualitative research technique, depth-interviewing was applied since this research also involved 'conducting intensive individual interviews when 'detailed information about a person's thoughts and behaviors' (Boyce & Neale, 2006, p. 3) were to be explored. Thirdly, the written reflection of the pre-service EFL teachers as self-report methods was also used as the data collection instrument since the "reflection is the process through which teachers become aware of the complexity of their work and are able to take actions which impact positively on this" (Pollard, et al., 2008, p. 476). Last of all, as the EFL pre-service teachers were required to write a critical essay paper on the evaluation of their attendance and contribution to teaching practicum as a whole process at the end of the term, their reflection was also used as the data in the research. However, as Ellis and Barkhuizen stated that "Because of the doubts that exist about validity and reliability of self-report methods, it is advisable to combine two or more self-report methods" (2016, 47), a combination of questionnaire, interviews and written reflections were used in order to provide more reliability and validity in this study.

3.4. Procedure

A research proposal, which included the whole procedure in detail in terms of its significance, purpose, methodology and tool of research, was presented to the dean of Necatibey Education Faculty of Balikesir University, Turkey in order to carry out this research. Soon after the proposal was considered and accepted, the researcher informed the participants about the procedure. Since this study involved four different data collection instruments such as background questionnaire, interview, reflection and essay paper on the research subject, each of the participants were informed about the fact that their rights would be protected in each process and that they would be also informed about the results of the research. The study used a four-step procedure: instruction, practice, data collection and data analysis.

3.4.1. Instruction

In the instruction part, the researcher introduced the whole procedure on teaching practicum giving some information on the theoretical aspects of teaching practicum and on how the teaching practicum could best be carried out. The researcher also made some

notifications on the preparation of the teaching practicum files as it was under the responsibility of pre-service EFL teachers in order to get a passing grade at the end of the process. After the introduction about the course content, the researcher made a clear explanation on the research and attempted to describe the whole process in detail. When the preservice EFL teachers gave their both oral and written consents on agreeing to be a part of the research, the researcher began her research legally.

3.4.2. Practice

In the practice part, after giving prerequisite information about the research, the preservice teachers were grouped in five-person groups, which means that there were five groups, each of which has a different time schedule in relation to their participation and contribution to the study. The timetable of the schedule, which showed each of the groups' participation weekly, was shared with the participants. However, they were also informed about if there had been something unusual about their experience at the teaching practicum course, they could reach the researcher by trying any type of social media accounts of her and her e-mail, even from her mobile phone. The reason why the researcher wanted the participants to share any kind of information by using any communication means was to make the participants feel secure and feel free as they were the main source of data. Accordingly, after a background questionnaire was employed in the first week, each group member visited and interviewed with the researcher weekly. Then, the participants handed their written reflections in fortnight intervals, which accounts for a seven-week written reflection in total within a 14 week of study. At the end of the process, the participants wrote a critical essay paper on the evaluation of their full attendance to teaching practicum courses as a whole process.

3.4.3. Data Collection

The data collection instruments including a background questionnaire, interviews, reflection and essay papers were used in order to ensure data triangulation, trustworthiness and validity of the data. The main purpose of using a background questionnaire was to obtain information about the participants' age and gender. Firstly, all participants were asked to fill a background questionnaire about their age, gender and educational background. Secondly, the data were collected through interviews. In the interviews, the pre-service EFL teachers were grouped in pairs and the researcher carried out the process by presenting some open-ended questions to the participants such as how they felt when they were at the practice school; what made them feel happy, upset, stressful, disappointed, etc.; whether they became anxious or not when they were at some of the intervals of teaching process; if they were anxious, when, why and how they felt such feelings of discomfort; what the anxiety-provoking factors were and personally what they did to overcome these problems. It must be noted that the interviews were conducted individually in order not to make the other participants interact with each other and to guarantee freedom of thought. In order to triangulate the data collection, and to provide further validation to the collected data, semi-structured interviews were conducted with each of the participants. In the study, the data were collected through reflections of pre-service EFL teachers in each week since the reflection was considered as 'a process that can facilitate both learning and understanding and plays a central role in several recent models of teacher development' (Richards, 1998, p. 21). Moreover, "Reflection is viewed as the process of critical examination of experiences, a process that can lead to a better understanding of one's teaching practices and routines" (Richards & Farrell, 2005, p. 7). It must be noted that the reflection was purposefully chosen to collect data as it is functional in teacher education particularly in improving the teaching skills of English student teachers. And lastly, at the end of the teaching practicum, the participants were asked to write an essay paper which clearly elucidate the whole procedure

they had experienced and their performance on teaching practicum. They also commented on a number of interesting experiences that they had gained during teaching practicum.

3.4.4. Data Analysis

The data collected from each source, namely a background questionnaire, interviews, reflection and essay papers were analyzed individually and the statements on anxiety-provoking factors among pre-service EFL teachers were coded by using descriptive and content analyses. The statements of the participants were coded under certain themes as coding is the paramount activity of the analysis (Miles & Huberman, 1994) in a qualitative research. In the data analysis process, similar statements were classified into specific themes under such as cognitive, affective and socio-cultural factors. In other words, the data collected from each source were transferred into concept maps in which the statements were classified. After the comparison of the statements and numbers in concept maps and codification of themes, it was found that the data on anxiety provoking factors among pre-service teachers seemed similar. The percentages of these codifications were determined and the related tables were created. In other words, the data were combined and the numbers and frequencies were given, which can be seen in the following sections.
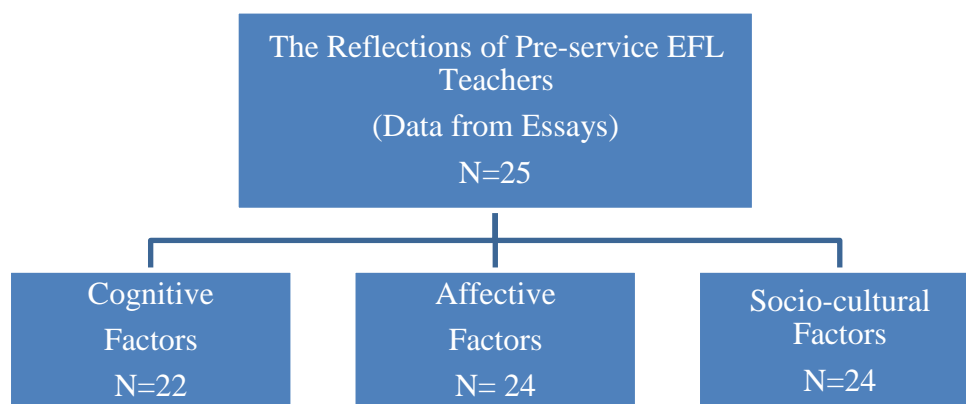
The Reflections of Pre-service EFL Teachers (Data from Essays) N=25

Cognitive Factors N=22

Affective Factors N= 24

Socio-cultural Factors N=24

Figure 1. Sample concept map

4. Findings

As this study attempted to explore what the pre-service teachers' opinions on teaching practicum were in general, a number of excerpts taken from preservice EFL teachers' reflections were given in the following part.

Among the 25 pre-service EFL teachers, 22 of them were of the opinion that teaching practicum was very important, effective and a well-designed course theoretically. They stated that they had a valuable experience on the practices of teaching itself. They also revealed that if the course was applied both theoretically and practically as it was indicated in the procedure, it would certainly be the best of the courses ever in English Language Teaching departments.

Among the positive reflections, one of the participants stated that "*Teaching practicum is a constructive and accommodating course, but when applied properly. However, it is also debatable to what extent it is applied appropriately. Teaching practicum is very important in*

the process of language teaching as it helps us gain tolerable amount of experience"
(Recorded Interview, 2).

Among 25 pre-service ELT teachers 18 of them stated negative opinions on teaching practicum as the practice time of the year and duration were not suitable for the 4th class students because of their KPSS preparation; as no information was given by the practice teacher about in-class discourse and attitude of the teachers, and also about the procedure of school administration; as being introduced by practice teachers as 'stajer öğretmen' (novice teacher); as the students do not show respect to them; as practice teachers have hardly any idea on how to carry out this course; as the practice schools are far away from their houses and faculty of education; as the pre-service EFL teachers are not paid by the government.

One of the participants stated that "*My point of view on teaching practicum is completely negative as it involves many-sided deficiencies, one of which is related with the lack of coordination between practice teachers and teacher trainers. Moreover, as the teaching practice course is in the last year in the faculty of education, I do not believe that it reaches its target. I can easily show the increased responsibility on us because of KPSS exam as the reason of this*" (Recorded Interview, 11).

This study also attempted to explore whether the pre-service EFL teachers felt teaching anxiety before and during teaching practicum or not, a number of excerpts taken from pre-service EFL teachers' reflections were given in the following part. However, the responses of the participants were evaluated as 'before the teaching practicum' and 'during the teaching practicum'.

As for 'before the teaching practicum', among 25 pre-service EFL teachers, 20 of them stated that they felt anxious before teaching practicum. The main factor of the increase of their anxiety was that they had never had the chance to practise their English in a real classroom setting. They were also anxious before teaching practicum whether the practice teachers and directors of the practice schools would behave them kindly or not; whether the level of the students in the practice school would be sufficient enough to show their own teaching ability; whether they would be unable to teach or to remember any kind of rules in English language teaching; whether they would have some difficulty in establishing a good rapport between the students; whether they would be able to manage the class; whether the class would be effective; whether the process of getting used to teaching would be long; whether the practice teachers would be intolerant about attendance. In addition to all these, some of the negative experiences told by previous pre-service EFL teachers; the feelings of inability about a particular subject in English and obscurity of the process of teaching practicum were also the sources of anxiety among them.

One of the participants stated that "*I felt extremely anxious before the teaching practicum. Because I was afraid of being responsible for a whole class. As I did not have any experience on teaching in a real class setting, I was doubtful that I would not be able to transfer the knowledge I have already had. Furthermore, I thought that even the rest of the teachers at the practice school did not consider us as a prospective teacher and they never show respect to us. Then I felt discouraged as I could not find a way to identify myself*" (Recorded Interview, 18).

Among 25 pre-service EFL students, 5 of them stated that they did not feel anxious before teaching practicum, as they had the chance of practicing teaching at real classroom settings either at a private school as a tutor or at homes as a private teacher. And one of the participants stated that "*I did not feel anxious before the teaching practicum. Because we had teaching practice course in the first semester and I thought that teaching practice in the*

second semester would be much or less similar to the one in the first semester. And I have already had a chance to practice teaching at a private school" (Recorded Interview, 24).

As for 'during the teaching practicum', among 25 pre-service EFL students, 23 of them reported that they felt anxious in the process of teaching practicum, particularly when they were unable to attract the students' attention; when they were unprepared and unplanned; when something unexpected happened in the class; when an unpredicted question was directed; when confronted with the misdeeds and mischievousness of the students; when being unsure on whether the students understood the subject or not. Most of the pre-service EFL teachers revealed that they were anxious on making pronunciation mistakes, on making errors of any kind, on forgetting something important and on not having adequate information on the process of teaching practicum. And additionally, one of the participants stated that "*I felt anxious in the process of teaching practicum as the practice teacher introduced us all to the students as 'stajer öğrenci' not even as a teacher. So, all the students in the class were doubtful about our knowledge, competency and teaching performance. It made me feel anxious from the beginning to the end of the process*" (Recorded Interview, 6).

Among 25 pre-service EFL students, only 2 of them revealed that they did not feel anxious in the process of teaching practicum. And, one of the participants stated that "*I felt very comfortable in the class as I have already had teaching experience in my life*" (Recorded Interview, 5).

Moreover, among 25 pre-service EFL students, 20 of them stated that they felt anxious while they were practicing at a real class setting particularly while trying to take the attention of the whole class; while balancing the level of noise in the class; while trying to communicate with the students; while using the time allowed properly, while trying to motive the students; while pronouncing some certain words; while planning the course. Some of the pre-service EFL teachers were also anxious as they had the fear of being unable to make the students understand the activities and the materials prepared. Interestingly enough, a pre-service EFL teacher experienced anxiety because of length. As she was short as a person, she believed that the students did not take her seriously. She revealed that physical characteristics affect psychological situations of the individuals very much. In addition, one of the participants stated that "*To what extent I am good at teaching the subject does not make any sense in some particular situations. Experience and practice make perfect! If something unexpected happens in the class, I am stunned and shocked. For instance, at a time, one boy started to kick out at another boy during the course. At first, I did not know what to do. But after a while, I made up my mind and tried to control the boys and the class*" (Recorded Interview, 21).

Among 25 pre-service EFL teachers, 5 of them stated that they did not feel anxious while they were practicing at a real classroom setting as they were able to answer the students' questions, as they were experienced and as they had given the impression of a real teacher to the students. One of the participants reported that "*As my practice teacher has already had the authority on the students, I did not have to worry about the classroom management in the class while practicing. Moreover, I have been working as a tutor since the first year of my university education. Thus, I am much more relaxed than the rest of the classmates*" (Recorded Interview, 17).

This study lastly attempted to explore what the anxiety-provoking factors during teaching practicum were. Thus, the research concluded that the anxiety-provoking factors among the pre-service EFL teachers were classified into three as cognitive, affective and socio-cultural factors.

4.1. Cognitive Factors

The following table indicated the anxiety provoking factors among pre-service EFL teachers in relation to cognitive factors.

Table 1. *Anxiety Provoking Factors among Pre-Service EFL Teachers in relation to Cognitive Factors*

Cognitive Factors	f	%
1. Being unable to teach the subject properly	8	27.59
2. Being academically assessed by practice teachers	4	13.78
3. Making grammar and vocabulary mistakes	4	13.78
4. Using the allowed time unwisely and inefficiently	2	6.90
5. Using the inappropriate materials	2	6.90
6. Being unable to practise at a heterogeneous class	2	6.90
7. Being unable to use interactive board	2	6.90
8. Being unable to answer students' questions	2	6.90
9. Being unable to transfer the knowledge	2	6.90
10. Being unable to know the content of the course book	1	3.45
Total	29	100

In terms of cognitive factors, the results revealed that the pre-service EFL teachers experienced anxiety when they thought and believed that they were unable to teach the subject properly; they were academically assessed by practice teachers; they made grammar and vocabulary mistakes; they used the allowed time unwisely and inefficiently; they used the inappropriate materials; they were unable to practise at a heterogeneous class as they could not think properly what to do systematically; they were unable to use the interactive board; they were unable to transfer the knowledge; they were unable to answer students' questions; they were unable to know the content of course book.

4.2. Affective Factors

The following table indicated the anxiety provoking factors among pre-service EFL teachers in relation to affective factors.

Table 2. *Anxiety Provoking Factors among Pre-Service EFL Teachers in relation to Affective Factors*

Affective Factors	f	%
1. Fear of losing class management while teaching	9	27.28

2. Fear of being negatively evaluated by practice teacher	4	12.12
3. The intolerance/indifference of practice teachers	4	12.12
4. The observation of practice teachers	3	9.09
5. Difficulty of accepting a new identity as a teacher	3	9.09
6. Fear of speaking before the public	3	9.09
7. Low self-confidence/self-esteem	2	6.06
8. Fear of being unprepared for the class	2	6.06
9. The practice teachers' unwillingness to the course	1	3.03
10. The lack of concession between the coterie teachers	1	3.03
Total	33	100

In terms of affective factors, the results revealed that the pre-service EFL teachers experienced anxiety when felt that they lost class control; they would be negatively evaluated by the practice teachers; they were mistreated or neglected by the practice teachers who showed signs of intolerance and indifference; they were observed by the practice teachers; they had difficulty in accepting a new identity as a teacher and in speaking before the public; they had low self-confidence and self-esteem; they were unprepared for the class; they caught the signs of practice teachers' unwillingness to the course; they realized the lack of concession between the coterie teachers as they knew that their friends were treated differently as some of the practice teachers were strict, some were tolerant and some were indifferent etc.; they realized that coterie teachers presented competing attitudes among each other in order to be more successful or to be more reputable and respectable among the parents.

4.3. Socio-cultural factors

The following table indicated the anxiety provoking factors among pre-service EFL teachers in relation to socio-cultural factors.

Table 3. *Anxiety Provoking Factors among Pre-Service EFL Teachers in relation to Socio-Cultural Factors*

Socio-cultural Factors	f	%
1. Negative attitudes of the students	6	22.22
2. The crowded classroom settings	4	14.82
3. The ecocultural setting of the schools	3	11.11
4. The excessive noise in class settings	3	11.11
5. The indifference of the students	3	11.11

6. The separation of the students as girls and boys	3	11.11
7. Difficulty in organizing group/ pair work	2	7.41
8. Difficulty in communicating with the coterie teachers	2	7.41
9. Being under surveillance by the other teachers	1	3.70
Total	27	100

In terms of socio-cultural factors, the results revealed that the pre-service EFL teachers experienced anxiety when felt and thought that the students in the class behaved them negatively; they had to overcome crowded classroom settings; they did not like the ecocultural settings of the schools; they could not cope with excessive noise; they came across with the indifference attitudes of the students; they could not discriminate the gender roles of the students; they had difficulty in organizing group/pair work; they had to communicate with the coterie teachers in the teachers' room; they felt they were under surveillance by the other teachers in the teachers' room.

In sum, the findings of this study revealed that the teaching practicum was one of the most significant courses in ELT departments according to pre-service EFL teachers. Whereas most of the pre-service (N= 22) teachers revealed their positive reflections on teaching practicum, some of them (N=18) stated their negative reflections. The findings also indicated that the participants experienced teaching anxiety before and during teaching practicum. While most of the participants (N=20) felt anxious before the teaching practicum, nearly all participants (N=23) were anxious during the teaching practicum. As the findings suggested a novel classification on the anxiety-provoking factors among pre-service EFL teachers, namely a kind of holistic classification rather than an atomistic one, this study concluded that anxiety-provoking factors can be classified into cognitive, affective and socio-cultural factors.

5. Discussion and Conclusions

Since the existing research has provided very limited insights into FLTA from the teachers' or pre-service teachers' perspectives, the research on FLTA remains rather vague and it requires further studies from a wide variety of perspectives and approaches as well. In other words, as an area of research to be addressed and explored thoroughly, FLTA deserves the attention of the researchers as FLTA has not been defined properly and it has been used in many research papers in lieu of FLA. Thus, in this study, firstly FLTA as an ill-defined concept is attempted to (re)define in language teaching. *Foreign language teaching anxiety* (FLTA), a kind of precisely *situation-specific anxiety* as Ellis (1994) labelled, can be defined as a psychological discomfort or apprehension which is unique for pre-service teachers, language teachers, instructors and practitioners who are responsible for teaching learners how to learn a foreign language effectively and professionally at a real classroom setting by using all skills and subskills required for teaching the target language. In other words, FLTA, which stems from cognitive, affective and socio-cultural factors, can also be defined as the feeling, believing and thinking of being inadequate in teaching and practising the foreign language as well as of being incompetent in performing the language acceptably with the real students at a real teaching context.

In this study, as the interview results clearly indicated, the accounts from the participants seemed to verify some of the findings offered by previous research on FLTA. For instance,

this study was similar to that of Merç (2004) as anxiety was among the most frequently reported problems among pre-service EFL teachers during teaching practicum. It was also similar to those of Gardner and Leak (1994) and Aydin (2016) as both of the studies revealed teaching experience and being inexperience was among the anxiety provoking factors. Similarly, it also corresponded with the study of Kim and Kim (2004) as some of the results such as being asked unexpected questions, being unable to control the class and being observed in the class showed similarity with the results of this study. Moreover, this study also produced almost identical results with those of Öztürk (2016) as teaching a specific skill in which the pre-service teachers feel incompetent can be a source of anxiety. It was also similar to that of Na (2007) as the results of the both studies indicated communication apprehension and fear of evaluation as anxiety-provoking factors. Moreover, it was comparable to those of Fletcher (1997) and Hancock (2001) studies as the results concluded in both studies that anxiety generates an adverse effect on performing foreign language. As a last point, this study also suggested low-self confidence and self-esteem were factors similar to those found by Yoon (2012).

Contrastively, the result of this study revealed some different results from those found in the literature. For instance, when the literature was reviewed, among the factors that caused teaching anxiety there existed personality or identity-based anxiety (Anandari, 2015; Aydin, 2016; Stroud & Wee, 2006). However, in this study, none of the participants mentioned personality or identity-based anxiety as an obstacle on FLTA. This study was also different from that of Thomson and Lee (2013) as the fear of ambiguity was not found to be as an anxiety provoking factor. Moreover, this study did not correspond to the study of Tum (2015) as the pre-service teachers did not mention that they refrained from using the target language and language intensive teaching practices when their anxiety level increased.

Speaking specifically, this research on the sources of teaching anxiety mainly bases its premises on holistic classification rather than atomistic one. Thus, it attempts to suggest a more general classification on the sources of FLTA. Accordingly, the results obtained from the study indicated that pre-service EFL teachers experienced anxiety before and during teaching practicum and that the sources of anxiety-provoking factors are classified under cognitive, affective and socio-cultural factors.

Some practical recommendations can be noted. FLTA can be included in the curriculum by adopting some concrete resolutions to overcome anxiety-provoking factors. In other words, pre-service EFL teachers should be trained cognitively, affectively and socio-culturally on how to overcome FLTA. Only under these circumstances, the pre-service EFL teachers may be trained and familiarized with the FLTA in order to carry out teaching practicum effectively. As teaching practicum is the last phase of being a teacher, the pre-service EFL teachers should be informed on many sources of teaching aspects either by the practice teachers or by the faculty members. Moreover, governmental precautions should be taken urgently and every practice teacher who is responsible for pre-service teachers should be trained pedagogically, administratively and theoretically. As this study was restricted to 25 pre-service EFL teachers at a state university and was confined to the qualitative data obtained from interviews, reflections and essay papers, it must be noted that the data obtained from this study was rather limited.

the Association of Istanbul Science and Academicians in Istanbul, on 22/23 December, 2017. Moreover, I would like to show my gratitude to the anonymous reviewers who provided insights and expertise that greatly assisted the research.

References

Alpert, R., & Haber, R. N. (1960). Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology*, *10*, 207-215.

Anandari, C. L. (2015). Indonesian EFL students' anxiety in speech production: Possible causes and remedy, *TEFLIN Journal: A publication on the teaching and learning of English*, *26*(1), 1-16.

Aslrasouli, M., & Vahid, M. S. P. (2014). An investigation of teaching anxiety among novice and experienced Iranian EFL teachers across gender. *Procedia-Social and Behavioral Sciences*, *98*, 304-313.

Aydin, S. (2016). A qualitative research on foreign language teaching anxiety. *The Qualitative Report,* *21*(4), 629-642. Retrieved from http://nsuworks.nova.edu/tqr/vol21/iss4/2

Baltacı, H.S. Foreign language teaching anxiety of language teachers pre-service vs. in-service, 1ˢᵗ International Black Sea Conference on Language and Language Education, September 22-23, 2017, Ondokuz Mayıs University, Samsun, 291-296.

Boyce, C., & Neale, P. (2006). *Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input.* Pathfinder International Tool Series, Monitoring and Evaluation-2. Retrieved from http://compositionawebb.pbworks.com/f/conducting+in+depth+interviews.pdf

Chastain, K. (1975) Affective and ability factors in second language acquisition. *Language Learning*, *28*, 55-68.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). California: Sage Publications.

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford: Oxford University Press.

Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.

Ellis, R. & Barkhuizen, G. (2016). *Analysing learner language*. Oxford & New York: Oxford University Press.

Fish, T. A., & Fraser, I. H. (2001). Exposing the iceberg of teaching anxiety: a survey of faculty at three New Brunswick Universities. *Electronic Journal of the American Association of Behavioral and Social Sciences*, *4*.

Fletcher, C. (1997) A study of state, trait, and test anxiety, and their relationship to assessment center performance. *Journal of Social Behavior & Personality, 12*(5), 205- 214.

Freud, S. (1924). *Collected papers*. (Vol. 1). London: Hogarth Press.

Freud, S. (1936). *The problem of anxiety.* New York: W.W. Norton.

Gardner, L. E., & Leak, G. K. (1994). Characteristics and correlates of teaching anxiety among college psychology teachers. *Teaching of Psychology*, *21*(1), 28-32.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitude and motivation*. London: Edward Arnold.

Hancock, D. R. (2001) Effects of test anxiety and evaluative threat on students' achievement and motivation. *The Journal of Educational Research*, 94, 284-290.

Horwitz, E. (1996). Even teachers get the blues: Recognizing alleviating language teachers' feeling of foreign language anxiety. *Foreign Language Annals*, *29*(3), 365-372.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal, 70*(2), 125-132.

İpek, H. (2007). Foreign language teaching anxiety. Eskişehir: T.C. Anadolu Üniversitesi Yayınları, No: 1730.

İpek, H. (2016). A qualitative study on foreign language teaching anxiety. *Journal of Qualitative Research in Education*, *4*(3), 92-105. doi:10.14689/issn.2148-2624.1.4c3s5m

Kim, S. Y., & Kim, J. H. (2004). When the learner becomes a teacher: Foreign language teaching anxiety as an occupational hazard. *English Teaching*, *59*(1), 165-186.

Lim, H. Y. (2009). Culture, attributions, and language anxiety. *Applied Language Learning*, *19*, 29-52.

Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *The Modern Language Journal*, *92*(1), 71-86.

MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second-language learning: Toward a theoretical clarification. *Language learning*, *39*(2), 251-275.

MacIntyre, P. D., & Gardner, R. C. (1994). The subtle effects of language anxiety on cognitive processing in the second language. *Language learning*, *44*(2), 283-305.

Merç, A. (2004). *Reflections of pre-service EFL teachers throughout their teaching practicum. What has been good? What has gone wrong? What has changed?* (Unpublished master's thesis). Anadolu University, Eskişehir.

Merç, A. (2011). Sources of foreign language student teacher anxiety: A qualitative inquiry. *Turkish Online Journal of Qualitative Inquiry*, *2*(4), 80-94.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis.* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Na, Z. (2007). A study of high school students' English learning anxiety. *The Asian EFL Journal*, *9*(3), 22-34.

Numrich, C. (1996). On becoming a language teacher: Insights from diary studies. *Tesol Quarterly*, *30*(1), 131-151.

Oxford, R. L. (1999). Anxiety and the language learner: new insights. In Jane Arnold (Ed.), *Affect in language learning* (pp. 58-67). Cambridge: Cambridge University Press.

Öztürk, G. (2016). Foreign language teaching anxiety among non-native teachers of English: A sample from Turkey. *Sakarya University Journal of Education*, *6*(3), 54-70.

Pollard, A., Anderson, J., Maddock, M., Swaffield, S., Warin, J., & Warwick, P. (2008). *Reflective teaching* (3rd ed.). London: Continuum.

Richards, J. C. (1998). Beyond training (1st ed.). Cambridge: Cambridge University Press.

Richards, J. C., & Farrell, T. S. C. (2005). *Professional development for language teachers: Strategies for teacher learning*. USA: Cambridge University Press.

Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning, 28*(1), 129-142.

Spielberger, C. D., Gorsuch, R. L., Lushene, R., Vagg, P., & Jacobs, G. A. (1983). *Manual for the state-trait anxiety inventory*. Palo Alto, Calif.: Consulting Psychologists Press.

Stroud, C., & Wee, L. (2006). Anxiety and identity in the language classroom. *RELC Journal*, *37*(3), 299-307.

Thompson, A. S., & Lee, J. (2013). Anxiety and EFL: Does multilingualism matter? *International Journal of Bilingual Education and Bilingualism*, *16*(6), 730-749.

Trang, T. T. T., Baldauf, R. B., & Moni, K. (2013). Foreign language anxiety: Understanding its status and insiders' awareness and attitudes. *Tesol Quarterly*, *47*(2), 216-243.

Trylong, V.L. (1987). *Aptitude, attitudes, and anxiety: A study of their relationships to achievement in the foreign language classroom* (Unpublished doctoral dissertation). Purdue University, West Lafayette, IN.

Tum, D. O. (2012). Feelings of language anxiety amongst non-native student teachers. *Procedia: Social and Behavioral Sciences, 47*, 2055-2059. doi:10.1016/j.sbspro.2012.06.948

Tum, D. O. (2015). Foreign language anxiety's forgotten study: The case of the anxious preservice teacher. *Tesol Quarterly*, *49*(4), 627-658.

Yin, R. K. (2010). *Qualitative research from start to finish.* NY: SAGE Publications, Inc.

Yoon, T. (2012). Teaching English through English: Exploring anxiety in non-native pre-service ESL teachers. *Theory and Practice in Language Studies*, *2*(6), 1099-1107. doi:10.4304/tpls.2.6.1099-1107

Young, D. J. (1991). Creating a low-anxiety classroom environment: What does the anxiety research suggest? *Modern Language Journal*, *75*, 426-439.

Young, D. J. (1992). Language anxiety from the foreign language specialist's perspective: Interviews with Krashen, Omaggio Hadley, Terrell, and Rardin. *Foreign Language Annals, 25,* 157–172. http://dx.doi.org/10.1111/ an.1992.25. issue-2

MEASUREMENT INVARIANCE: CONCEPT AND IMPLEMENTATION

Güçlü Şekercioğlu [iD]

Akdeniz University, Turkey

guclus@akdeniz.edu.tr

Güçlü Şekercioğlu has had his Ph.D. in Measurement and Evaluation from Ankara University, Turkey. He is a teaching staff at Akdeniz University, Education Faculty. He conducts research in multivariate statistics, measurement invariance, differential item functioning and psychological measurement in culture context.

MEASUREMENT INVARIANCE: CONCEPT AND IMPLEMENTATION

Güçlü Şekercioğlu

guclus@akdeniz.edu.tr

Abstract

An empirical evidence for independent samples of a population regarding measurement invariance implies that factor structure of a measurement tool is equal across these samples; in other words, it measures the intended psychological trait within the same structure. In this case, the evidence of construct validity would be strengthened within the frame of the scores obtained from the tool. When measurement invariance is not supported, the researchers should consider the possibility of the different factor designs for each group. Ignoring such a situation brings forward the probability about differentiation of the trait(s) measured by measurement tool for that/those group(s), so it causes to suspect the validity of the scores obtained from the tool. The aim of this study is to examine measurement invariance in the context of the conceptual foundations of multi-group confirmatory factor analysis, and discuss the subject through the results from two hypothetical data set that one supports measurement invariance, but the other does not. As a result of analysis performed in this direction, it is determined that the five-factor design derived from the first data set is equal across the groups in the majors of science, health, and social science. It is also concluded that the three-factor design obtained from the secondary data set is not equal for female and male groups. Besides, the exploratory factor analysis performed for female and male groups separately shows that the three-factor design of the tool is valid for females, but the number of factors was four in males. When the factor design for male group is examined, it is determined that the three items in the second factor separate significantly. That leads to the conclusion that it is crucial to test measurement invariance in studies regarding the determination of the psychometric properties of the tool.

Keywords: measurement invariance, equality of factor structures, multi-group confirmatory factor analysis, structural equation modeling

1. Introduction

The major problem in behavioural and educational science studies, which aim developing the psychological measurement tools, cultural adaptation of a tool developed in another culture, using the tool for a different purpose or for a different sample, is to demonstrate the validity of the empirical evidence on the psychometric properties of the tool. In this direction the researchers, within the framework of these fundamental problems, are obliged to question whether the tool measures the trait(s) what it intends to measure properly and precisely. Further examination related to psychometric properties of measurement tool and all other analyses based on the scores obtained from the measurement tool (ANOVA, regression, etc.) has been carried out after the validity of the evidence put forward and decision are taken in this direction. According to Nunnally and Bernstein (1994), the validity of each usage must be documented by empirical evidence even though a measurement result may be valid for more than one purpose. Therefore, the test authors and users should not assume that the validity of evidence cannot change (Crocker and Algina, 1986).

One of the most important dimensions of the validity of scores obtained from psychological measurement tools is the construct validity. In the report of testing standards published in 1954 it was discussed that the concept of validity, actually all types of validity should be assembled under the roof of construct validity (Cronbach and Meehl, 1955; Jonson and Plake, 1998; Urbina, 2004; Westen and Rosenthal, 2005). Similarly, Kline (2000) states that the construct validity includes other approaches as well, thus all types of validity are related to the assessment of construct validity. The factor analysis is one of the most commonly used techniques in the studies which aim to determine the psychometric properties of a measurement tool in behavioural and educational science, in order to obtain evidence of construct validity. According to Büyüköztürk (2002, 2014) the factor analysis is a multivariate statistics, which aims to find and explore conceptually meaningful fewer new variables (factor, component) by bringing a large number of inter-related variables together. The factor analysis can be considered under two headings, which are exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) discussed under the concept of structural equation modelling (SEM).

CFA, which is based on testing of theories about the latent variables, and used in advanced research, is a very sophisticated technique (Ullman, 2001). In this analysis, a predefined and constrained construct is tested whether it is confirmed as a model. It is also occasionally used to mean the confirmation of the theoretical structure (Maruyama, 1998). In this context, the determination of the construct validity for CFA is emphasized as a very powerful method (Floyd and Wideman, 1995; Kline, 2005; Stapleton, 1997).

Multi-group confirmatory factor analysis (multi-group CFA) is also a spesific practice area in CFA. This analysis enables to test the equality of structural parameters for more than one group simultaneously. In this context, the assessment of equality between the groups in terms of factor structure is also termed as *measurement invariance*. Additionally, examining the fitness of structure brings about the concept of testing population heterogeneity. It is possible to encounter different terms for different tests of measurement invariance tests including equality test of factor structures, metric and factorial invariance in the literature (Brown, 2006).

Nowadays, interest of the researchers in social sciences towards measurement invariance is gradually increasing. In a plain defination, measurement invariance is the description of whether the structures of measurement tool are equal across individuals from different groups. This concept has a critical importance in comparing groups. When measurement invariance is not supported between the groups, it is not possible to interpret the findings that reveal differences concerning these groups. If the researcher does not have the evidence for measurement invariance, then the existence of different psychometric responses for scale items more than one group cannot be known. Measurement invariance analyses are used in intercultural comparison for groups speaking different languages in a culture, scale adaptation studies, the comparison of groups with different academic achievement, the comparison of employee groups in different areas of industry, comparisons based on gender and are also used to compare a control group and an experimental group in empirical research (Cheung and Rensvold, 2002). The frequently asked question on the use of psychological measurement tools is whether the factor design ensued as a result of factor analysis of the measurement tools valid for groups, which differentiated at such a level that may impact the measurement process concerning the ethnic characteristics, age or the way they respond to the items. In fact, the fundamental issue here is whether the measurement tool measures the same structural properties for different groups or not. When the factor structure is not equal across groups, naturally it is not possible to make meaningful comparisons between groups based on the factor scores. On the contrary, when measurement equivalence is supported

empirically, it is concluded that the group differences are completely reflected in terms of latent traits evaluated by factors. In this concept, the studies, which aim at determining equality of the measurement tools' factor structure, are becoming more and more substantial because cultural, developmental and contextual impacts related to the psychological structural traits have become well-known by the researchers recently (Floyd and Widaman, 1995). In addition to social science studies, using multi-group CFA becomes increasingly common in other majors such as psychology, education, management and organization, marketing, and communication, especially ones which carried out studies based on cross-cultural comparisons.

As Jöreskog, Sörbom and Toit (2000) claim, the factor structure of developed or adapted scale based on fundamental data set obtained from different groups or samples can be tested whether it is equal for more than one group or not concerning the national, territorial, regional, cultural or socio-economic status of the groups. It is highly functional to test the equality of factor structures for a scale or different numbers of items for more than a group. Thus, factors or structural relationships can be tested simultaneously whether they are equal across different samples (Baumgartner and Steenkamp, 1998).

According to Marcoulides and Schumacker (1996) in multi-group CFA, the question of "*is each group measured under the same structure?*" is investigated and this examination is carried out within the framework of the measurement model defined in advance. Similarly, Kline (2005) stated that, the focus of multi-group CFA is to test whether measurement invariance is supported for different groups within the same latent variables. This concept is defined as invariance of the psychometric properties of a scale across groups in the context of modelling in the literature of psychometry.

Determining whether the measurement invariance is supported for different groups or not has a critical role in the development of psychometric properties of psychological measurement tools. That implies whether the items of the same structure and all structures can be used for the sub-groups of a population. Likewise, the subject of testing measurement invariance plays a crucial role in terms of defining the generalizability of psychological structure across groups with differinf variables such as different cultures, age groups and genders. The equality tests of latent means, which are included in the analysis group, shows similarity with the comparison of observed group averages through t-test and ANOVA (Brown, 2006).

According to Byrne (2006), the researchers often seek answers to any of the following five questions for evidence related to the multi-group equality: (i) do certain structures of the items on the measurement tool work equally across different groups? In other words, does the measurement model have a group equality? (ii) is the factor structure of the tool or theoretical structure measured by multiple scales equal for each level of the group? (iii) are the paths of the experimental structures equal across the groups? (iv) does the latent means in the model for a particular structure vary between groups? and (v) is the factor structure of a measurement tool equal for independent samples of the population? The author particularly emphasizes that there could be a cross-validation study in his last question. The analysis results reach the conclusion that if the factor structures are not equal between the groups, the validity of interpretations based on a comparison of scores for these groups decreases.

According to Brown (2006), the process steps below should be followed in the evaluation of the multi-group CFA and measurement invariance: (i) performing CFA for each group included in the analysis separately, (ii) testing the equality of structures simultaneously (factor loadings, factor correlations and error variances constant), (iii) testing the equality of the factor structures (factor loadings free; factor correlations and error variances constant),

(iv) testing the equality of factor structures and the error variances indicators (factor loadings and error variances free; factor correlations constant), (v) testing the equality errors variances of indicator (error variances free, factor loadings and factor correlations constant), (vi) testing the equality of factor variances, (vii) testing the equality of factor covariances (if more than one factor), and (viii) testing the equality of latent means. Hereunder, the first step is one of the multi-group CFA's assumption. The processes between the second and fifth steps are about testing measurement invariance, and the processes between the sixth and eighth steps are about testing the population heterogeneity.

1.1. Measurement Invariance Test and Models

Before computing the multi-group CFA, first of all, correlation or covariance matrix of the groups in the same sample is evaluated by comparing each other. In other words, before setting up the configural invariance model (Model 1), the establishment of the test equality of covariance matrices (Model 0) must be made. If the equality of covariance matrices is provided for each group $(\Sigma^g = \Sigma^{g'})$ the configural invariance model can be developed and tested. The equality of covariance structures of the groups should be discussed only after the null hypothesis (H_0) has been rejected. Subsequently, the models for other hypothesizes should be tested separately. The configural model derived from different groups should be defined in the same sample. Thus, the defined model for each group of multi-group analysis would be simultaneously tested. In this case, it is expected to see high fitness between correlation or covariance matrices of different groups (Brown, 2006; Byrne, 2006; Dunn, Everitt and Pickles, 1993; Vandenberg and Lenca, 2000). In general, the measurement invariance is tested with four basic models. These models are summarized in Table 1 (adapted from Cheung and Rensvold, 2002).

Table 1. *Measurement invariance models*

Models	Hypothesis	Hypothesis Name	Symbolic Statement	Process
1	H_{form}	Configural invariance	$\Lambda^{(1)}_{form} \therefore \Lambda^{(2)}_{form}$	Invariance is supported for all groups regarding construct and items. Factor loadings, factor correlations, and error variances are equal for all groups.
2	H_Λ	Weak Factorial Invariance (Metric Invariance)	$\Lambda^{(1)} \therefore \Lambda^{(2)}$	Invariance is supported for all groups regarding factor correlations and error variances. The factor loadings have been freed for groups.
3	H_λ	Strong Factorial Invariance (Scalar Invariance)	$\lambda^{(1)}_{ij} \therefore \Lambda^{(2)}_{ij}$	Invariance is supported for all groups regarding factor correlations. The factor loadings and error variances have been freed for groups.
4	$H_{\Lambda,\Theta(\delta)}$	Strict Factorial Invariance (Residual Variance Invariance)	$\Theta^{(1)}_\delta \therefore \Theta^{(2)}_\delta$	Invariance is supported for all groups regarding factor loadings and correlations. The error variances have been freed for groups.

1.1.1. Configural invariance (baseline model)

Developing a configural invariance (also known as baseline model) begins with identifying and testing the model, which was developed within the framework of a specific hypothesis for each group. In this context, the number of sub-scales in configural invariance model for each group (e.g. factors), the positions of the items (e.g. which factors the items are loaded) and correlations between sub-scales (e.g. setting such factors covariance) are determined. Secondly, the validity of the configural invariance model is tested separately for each group. Ideally, the model is expected to well fit and significant. However, the evidence, which shows a well fit, provides the information to the researcher that only the factor structure is similar but does not give any certain information about the equality of factors for each group. The evidence act as a design for invariance tests to be carried out subsequently. This model has two important functions: (i) the parameters are tested simultaneously for all groups, (ii) equal initial value is generated for the integration of configural invariance model

for testing (Byrne, 2008). Hence, the criterion, which will be obtained from further models to be tested, is occurred. In this model, invariance, regarding structure and items are supported for all groups (factor loadings, factor correlations and error variances are equal for all groups). When weak, strong or strict factorial invariance hypothesizes are rejected, the "factor structure is equal across all groups" hypothesis, which is developed within the framework of configural invariance, is accepted.

1.1.2. Weak factorial invariance (metric invariance)

In this model, the equality of factor loadings (λ), ($\Lambda^1 = \Lambda^2 = ... = \Lambda^G$) is tested for all groups. (Spini, 2003; Vandenberg and Lence, 2000). If the fit, which is obtained from weak factorial invariance test, is better than the fit of configural invariance, configural invariance model is rejected. In other words, it indicates that the equivalence is not supported. According to Byrne and Stewart (2006) although measurement units are identical for groups in terms of underlying factors (e.g. factor loads), it constitutes one of the constraints of this model because scaling (e.g. intercepts) is not identical. Therefore, Meredith (1993) describes this invariance level as *weak factorial invariance*. This invariance is tested,

$$M_g \cong \tau_g\hat{\tau}' + \hat{\Lambda}(\alpha_g \hat{\alpha}' + \hat{\Psi}_g)\hat{\Lambda}' + \hat{\Theta}_{\varepsilon g} = \hat{M}_g'$$

with this equation (Widaman and Reise, 1997).

1.1.3. Strong factorial invariance (scalar invariance)

It is tested whether the regression constant (τ) of observed variables on the latent variables is equal across groups ($\tau^1 = \tau^2 = ... = \tau^3$) (Schmitt and Kuljanin, 2008). In this model, there are a series of additional constraints described in weak factorial invariance. These additional constraints include the intercepts of the variables that are observed in the matrices τ_g. If estimations are problematic in terms of invariance on groups, subscript g on matrix τ is removed. In this case, invariance is tested,

$$M_g \cong \hat{\tau}\hat{\tau}' + \hat{\Lambda}(\alpha_g\hat{\alpha}' + \hat{\Psi}_g)\hat{\Lambda}' + \hat{\Theta}_{\varepsilon g} = \hat{M}_g'$$

with this equation (Widaman and Reise, 1997).

1.1.4. Strict factorial invariance (residual variance invariance)

In this last model of the measurement invariance, about error terms across the groups $H_{\Lambda\phi}$ model limits ($H_{\Lambda\phi\theta}$) model equally ($\theta^1 = \theta^2 = \cdots = \theta^G$). With the addition of this constraint, testing the hypothesis of equality of measurement errors becomes possible for independent samples of the population. If the error variances are equal, it means the items have equal reliability in terms of groups (Spini, 2003). Strict factorial invariance is also created through strong factorial invariance as it occurs in strong factorial invariance created through the weak factorial model constraints. These additional constraints are defined as strict factorial invariance, which contains unique factorial invariance in $\hat{\Theta}_{\varepsilon g}$ matrix or measurement errors. This invariance is tested,

$$M_g \cong \hat{\tau}\hat{\tau}' + \hat{\Lambda}(\alpha_g \hat{\alpha}_g' + \hat{\Psi}_g)\hat{\Lambda}' + \hat{\Theta}_{\varepsilon} = \hat{M}_g'$$

with this equation (Widaman and Reise, 1997).

It should be noted that there are varrious classification in the related literature. Therefore, it is worth to consider following aspects suggested by Meredith (1993) and Dimitrov (2010), in the testing process of the equality of factor structures across groups, metric invariance is the general name of weak factorial invariance, strong factorial (scale invariance) and strict factorial invariance (invariance of error variance) models. However, there are some research

in literature that discuss the weak factorial invariance with the term of metric invariance. (Gregorich, 2006; Meade, Michels and Lautenschlager, 2007; Schmitt and Kuljanin, 2008; Spini, 2003; Vandenberg and Lance, 2000; Wu, Li and Zumbo, 2007). Besides, Cheung and Rensvold (2002) used the terms *metric invariance on construct-level* for weak factorial invariance, *item-level metric invariance* for strong factorial invariance and *error variance invariance* for strict factorial invariance.

Multi-group CFA for measurement invariance can be computed with such software statistical programs like LISREL, Amos, SAS/STAT, Mplus and EQS. The analysis starts with the creation of separate covariance matrices for the levels of the groups. It can be carried out by typing the syntax analysis in LISREL program or by following the instructions prescribed by the program (Toit and Toit, 2001). Measurement invariance is carried out in four models. The syntax samples of these models are named as EX10A.SPL, EX10B.SPL, EX10C.SPL, and EX10D.SPL in LISREL program. In the first model (Model 1), also known as configural invariance model, factor loads of structure(s), correlations and error variance are assumed to be equal and the analysis is run in this regard. The configural invariance model, which is a fundamental model for the equality of factor structure, is developed with the hypothesis that factor structures are equal (H_0=*There is no difference between factor structures*). In order to make comparisons with model defined in the analysis, a second alternative model named as weak factorial invariance model (Model 2) is analysed. In the weak factor invariance model, freeing the factor loads for each group, keeping the factor correlations and error variances constant are discussed. In the third alternative model strong factorial invariance (Model 3) factor loads and error variances for each level of the group are released, factor correlations are kept constant. The last and fourth model of measurement invariance is strict factor invariance model (Model 4). In this model while error variances are released, factor loads and factor correlations are kept constant (Byrne, 2010; Jöreskog and Sörbom, 1993; Toit and Toit, 2001).

1.2. Model Comparisons in the Decision of Measurement Invariance

In multi-group CFA invariance test, constrained and unconstrained model are compared. In terms of availability of different values for each group in constrained model, model parameters (e.g. factor loads) are not constrained in this model. The parameters have the same value for all groups in constrained model. When the fit of unconstrained model is better than the constrained one, it implies that constrained model is incorrect. In other words, if the unconstrained model fits better when the constrained parameters are released, they are allowed to get different values for each group, and the constrained model developed within the invariance hypothesis framework is rejected (Cheung and Rensvold, 2000).

For comparisons of models with multi-group CFA in the studies in which measurement invariance is tested, it can be said that there are two widely used approaches in literature. The first one is the comparison between configural invariance model developed with the hypothesis that there is no difference in factor structure for each group and alternative models (e.g. weak factorial invariance, strong factorial invariance and strict factorial invariance models). Hereunder, the first comparison is made between configural invariance model and weak factorial invariance model (model 1 and 2), the second is between configural invariance model and strong factorial invariance model (model 1 and 3), and the last one is between configural invariance model and strict factorial invariance model (model 1 and 4). According to this approach, in the case of equality of fit between any alternative model and configural invariance or in the event of deterioration, the configural invariance model developed with the hypothesis that there is no difference in factor structure for each group is accepted. On the other hand, if the alternative model indexes differ from the configural invariance indexes

significantly (in favour of alternative models), H_0 hypothesis is rejected. In this case, the equality of factor structure and thus, the measurement invariance cannot be supported (Byrne, 2010; Jöreskog and Sörbom, 1993; Toit and Toit, 2001). In the second approach, the comparisons are performed by following stepwise process. According to this, the analysis starts with less limited models and then the models are assessed by using nested χ^2 method (Brown, 2006). Accordingly, in comparison to nested models; $H_{form} > H_\Lambda > H_\lambda > H_{\Lambda,\Theta(\delta)}$ is used as base. In other words, comparisons are made between configural invariance model and weak factorial model (model 1 and 2), weak factorial invariance model and strong factorial invariance model (model 2 and 3), strong factorial invariance and strict factorial invariance (model 3 and 4). According to Van de Vijyer and Leung (1997) if the fit of nested models is equal, more constrained model is frequently accepted. If this is not the case, the equality hypothesis is rejected (as cited in Spini, 2003). Cheung and Rensvold (2002) also suggest another comparison containing only one difference from the first approach. Although the first two comparison is the same, the authors suggest a comparison between weak factorial invariance and strict factorial invariance (model 2 and 4).

1.3. Decision Making of Measurement Invariance

While deciding whether the factor structures are equal for each group, the significance level of χ^2 matrix is required and the level should above .05 value, in other words, a non-significance value p is expected. This situation means that the covariance matrix of each of the defined groups do not differ significantly, thereby measurement invariance is supported. According to Jöreskog and Sörbom (1993), examples of acceptability of fit indices provided in Table 2 might be used for decision.

Table 2. *Acceptance of equality of factor structure in multi-group confirmatory factor analysis*

Problem	χ^2	df	p value	Decision
A	38.08	10	0.000	Reject
B	1.52	2	0.468	Accept
C	8.77	4	0.067	Accept
D	21.55	8	0.006	Reject
E	38.22	11	0.000	Reject

As seen in Table 2, models A, D, and E in which significance value p is a problem, are rejected whereas problem B and C are accepted. The criteria determined in the developing first years of multi-group CFA have been questioned over time. χ^2 has a possibility to increase its significance value if the number of samples increases, therefore, alternative models are investigated whether to accept the fit of factor structures within the model framework or not to assess the fit between covariance matrices. Among these, firstly, the value of χ^2 and degree of freedom should be compared. In this regard the χ^2 value obtained from the more constrained model, χ^2 value from less constrained model and the "delta" value (delta means the difference and its symbols is Δ) which is between the degree of freedom are calculated. $\Delta\chi^2$ and Δdf values are determined with this calculation. The significance level of χ^2 value obtained from this determination, is controlled in the level of p<.01 or p<.05 by comparing the critical values in the distribution table of χ^2 (Byrne, 2010; Jöreskog, 1971; Kline, 2005; Lee and Leung, 1982; Steiger, 2007; Van den Bergh and Van Ranst, 1998). In this case, H_0 and H_1 hypotheses can be developed in the following format:

H_0: There is no significant difference between the more constrained model and less constrained models in terms of fit.

> *H₁: There is a significant difference between the more constrained model and less constrained models in terms of fit.*

In this respect, if $\Delta\chi^2$, which is calculated on the basis of χ^2 differences in a particular Δdf level, is less than critical table values, H_0 is accepted. In other words, there is no significant difference between two models in terms of fit, therefore, the researcher can make a decision about measurement invariance based on $\Delta\chi^2$. On the other hand, if $\Delta\chi^2$, which is calculated on the basis of χ^2 differences in a particular Δdf level, is more than critical table values, H_0 is rejected. Thus, there is a significant difference between two models in terms of fit, and if this difference is in favour of the alternative hypothesis, the researcher can assume that measurement invariance is not provided on the basis of $\Delta\chi^2$.

In many studies in which analysis of SEM concept is applied, the distribution(s) may be remote from normal within certain tolerances. In the absence of normality in large samples, χ^2 value (S-Bχ^2) obtained from Satorra-Bentler correction produces close values to the χ^2 that is produced when the number of people in the sample and the distribution of the produces is normal. S-Bχ^2 is a rather reliable statistical test used to evaluate covariance structure models in various distributions and sample sizes (Byrne, 2006; Everitt and Howell, 2005). As in the other SEM analyses, such as multi-group CFA, which is carried out to obtain evidence of measurement invariance, S-Bχ^2 can only be calculated if the distribution of each group is far from the normal distribution. In multi-group CFA, which is carried out with the maximum likelihood method, T_s value should be calculated for S-Bχ^2 scaled difference in terms of evidence of measurement invariance between nested models. T_s is calculated

$$T_s = (T_0 - T_1) / c_d$$

with this equation. T_0 is the normal maximum likelihood χ^2 value for nested model, T_1 is the normal maximum likelihood χ^2 value for comparison (less constrained model) model, and c_d is the degree of difference test correction. c_d is calculated

$$c_d = [(d_0 * c_0) - (d_1 * c_1)] / (d_0 - d_1)$$

with this equation. d_0 is the degree of freedom of nested model, d_1 is the degree of freedom of comparison model, c_0 is the correction degree of nested model, and c_1 is the correction degree of comparison model. c_0 and c_1 are calculated

$$c_0 = T_0 / T_0^* \text{ and } c_1 = T_1 / T_1^*$$

with this equation. T_0^* is S-Bχ^2 value of nested model, on the other hand T_1^* is S-Bχ^2 value of comparison model. By comparing T_s, which is calculated for S-Bχ^2 difference degrees, with the critical values in χ^2 distribution table, it can be determined whether measurement invariance is supported (Brown, 2006; Satorra and Bentler, 2011).

Recently, it is widely used as an alternative to utilize from fit indices as well as to evaluate the χ^2 differences among the models in many research due to a large number of n. According to Cheung and Rensvold (2002) it is inadvisable to reject null hypothesis in case of obtaining an insignificant χ^2 value. χ^2 is statistically sensitive test for large samples, however, it is not a practical test for model fit. In such case, alternative fit indices should be offered for χ^2. The comparative fit indices (CFI, NNFI / TLI, RMSEA etc.) are among the most frequently recommended ones. Within this framework, it is observed that many goodness of fit indices are commonly used together to evaluate general fit of the model and to report it. ΔGFI's are used as an alternative for χ^2 in multi-group CFA which is performed to determine whether the factor structures are equal or not. As in χ^2, the configural model whose factor loads, factor correlation and error variance are released in covariance matrices of groups, in other

words, the model which is developed with the hypothesis that factor structures are equal, is the basic model like in alternative fit indices. For the evidence of measurement invariance, the differences between models can be evaluated with the comparison of indices such as ΔRMSEA, ΔCFI, ΔGamma Hat, ΔMc, ΔIFI, ΔAIC, ΔEVCI, ΔNFI, ΔTLI, and ΔSRMR.

The fit values are expected to become better for the equality of factor structures when the parameters like factor loads and error variance in covariance matrices of the group are released together or one by one. With this regard, the differences are evaluated by comparing the indices (e.g. ΔSRMR, ΔCFI and ΔRMSEA) between configural model and other alternative models or nested models. The configural model set up with the hypothesis that there is no significant difference between factor structures of each group is accepted if the fit indices of alternative models are lower than the ones in configural model. On the other hand, if the fit indices of other alternative model are higher than the ones in configural model or nested model, the fit across models is evaluated whether it differs significantly or not.

Cheung and Rensvold (2000; 2002) suggested cut-off points for ΔCFI significance level between modes in terms of measurement invariance after carrying out a study by using Monte Carlo method. Hereunder, when ΔCFI≥−.01 is provided, then configural invariance model is accepted. In contrast to this situation, if ΔCFI is between −.01 and −.02, there will be increasing doubt about invariance. If it is more than -.02 it can be said that the difference between constrained and unconstrained model will increase. In this situation, configural model is rejected. In this context, it is decided that the factor structures are not equal and therefore an alternative model should be sought. In addition, the critical values of ΔGamma hat and ΔMcDonald NFI are −.001 and −.02.

Chen (2007) suggested cut-off points for decision of measurement invariance by considering situations like sample size of CFI, RMSEA and SRMR indices and sample sizes in groups after carrying out a study, which aimed at testing sensitivity of goodness of fit indices through Monte Carlo method. Accordingly, it can be concluded that measurement invariance cannot be supported (case of noninvariance) if sample size is small (n<300), sample sizes of groups are not equal, pattern of variance is the same, there is a relationship like ΔCFI≤−.005, ΔRMSEA≥.010 or ΔSRMR≥.025 between groups in terms of weak factorial invariance test, and there is a relationship like ΔCFI≥-.005, ΔRMSEA≥.010 or ΔSRMR≥.005 between groups in terms of strong factorial invariance or strict factorial invariance. On the other hand, measurement invariance can be supported when sample size is sufficient (n>300), numbers of groups compared are equal, there is a relationship like ΔCFI≥−.010, ΔRMSEA≥.015 or ΔSRMR≥.030 between groups in terms of weak factorial invariance test, and there is a relationship like ΔCFI≥−.010, ΔRMSEA≥.015 between groups in terms of strong factorial invariance or strict factorial invariance.

An important point to be considered in assessing multi-group CFA comparison of the four basic models is type I and type II error possibilities. If the sample is small for a null hypothesis, type I error is likely occurred. However, if the sample is getting larger for alternative hypothesis, the difference of fit will be extended. In that case, type II error is likely occurred. For this reason, to minimize the type I and type II error possibility, the cut-off points should be determined efficiently (Hu and Bentler, 1998). In their maximum possibility $\Delta\chi^2$ studies which were performed with the indicators acting as continuous variables, French and Finch (2008) controlled type I error in the level of .01 and .05 between different models and sample numbers. The researchers revealed that the power of $\Delta\chi^2$ has a positive correlation with sample size, indicator number of each factors and factor number. Meade and Bauer (2007) also extrapolated the same results about $\Delta\chi^2$ (as cited in Sass,

Schmitt and Marsh, 2014). There is no doubt that this case is valid for other delta fit indices as well. However, this study didn't include detailed discussions on that subject because it was beyond the scope.

1.4. Objectives

Researchers of behavioural and educational sciences provide evidence through a sample on the validity of scores obtained from developed or adapted psychological measurement tools. After revealing the psychometric properties of the measurement tools, measurement process can be practiced on an independent group in the same sample and various decisions may be taken by means of obtained scores in the same or a different study. The fact that a measurement tool with confirmed factor structure for a sample may not be valid for the independent sub-groups in the relevant sample is a probability that researchers should pay attention. In such a case, the validity of decisions to be taken with scores obtained from groups will be suspicions. Within this scope, this research aims to discuss the conceptual basis of multi-group CFA in measurement invariance in terms of basic concepts and to introduce the subject through two hypothetic data set that one supports measurement invariance, but the other doesn't, for the researchers aiming to determine the psychometric properties of a measurement tool. Thereby, a new perspective will be introduced to the researchers aiming to determine the psychometric properties of measurement tools, suggestions about decisions to be taken for the tool without equalized factor structure will be asserted. Accordingly, the present study searches answers to the following research questions:

1. Is the five-factor structure of measurement tool 1 equal across the groups of science, health and social sciences?
2. Is the three-factor structure of measurement tool 2 equal across groups of males and females?

This research is limited to measurement invariance (measurement of configural invariance, weak factorial invariance, strong factorial invariance and strict factorial invariance). The heterogeneity of the population (factor variance invariance, factor covariance invariance and latent means invariance) is not included in the research.

2. Method

This study examines the method of multi-group CFA for the evidence concerning measurement invariance through two data set consisting of equal and unequal factor structure. Considering the findings of the study, the current study has the characteristics of correlational research concerning equality of factor structure for independent groups in two samples and due to the discussions on generation of construct validity evidences. The correlational studies analyse the relationship between two or more variables without intervening in these variables under any circumstances. These studies are the ones that are effective on revealing the relationships and determining the levels of relationships between variables and provides necessary cues for conducting high-level research on these relationships (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz and Demirel, 2012).

2.1. Research Data

The ready-made data was used in this study. They consist of two data set (equal and unequal factor structured) that the researchers collected them from his previous researches. The first hypothetic data set that measurement invariance is supported consist of 666 undergraduate students. When the distribution of the participants is examined based on scientific major, 32.28% (n=215) science, 31.83% (n=212) of health and 35.89% (n=239) of social science. The other hypothesis data set that measurement invariance is not supported

consist of 353 high school students. The distribution in terms of participants' gender is as follows, 62.32% (n=220) female, 37.68% (n=133) male.

2.2. Data Collection Tools

The study consists of two hypothetic data set, which are the subjects of measurement invariance analyses and the scores obtained from two different measurement tools. Some items were emitted from the tool in line with the results of EFA and CFA that were run on the data set collected from the participants. Moreover, the factor design differed for male participants in the second data collection tool whose factor design was not equal. The main purpose of this study is not to determine or discuss the psychometric properties of aforementioned tools. However the present study focuses on presenting the multi-group CFA in terms of measurement invariance through two hypothetic data set in which measurement invariance both was supported and was not, and creating a new view of validity for the researchers who aim to measure the psychometric properties of a measurement tool. Therefore, it is not appropriate to give the names of the tools and sub-scales in view of the probability that because they can form basis for further studies. For this reason, the data collection tools were mentioned as measurement tool 1 and measurement tool 2, and limited information about the psychometric properties of the tools was given because it was not wanted to reveal the tool.

Measurement tool 1 is a tool that consists of five sub-scales to measure an effective trait through using four point rating. In the original study, EFA and CFA were performed to determine psychometric properties of the tool in terms of gathering evidence about construct validity, concurrent validity was examined by comparing with a criterion score, to obtain reliability evidence for stability a test-retest method was run, and lastly to obtain reliability evidence for internal consistency, Cronbach alfa coefficients were calculated. In conclusion, it can be said that the scores obtained from measurement invariance tool 1 have a high level of validity.

This study starts with EFA to obtain construct validity evidence through hypothetic data set of measurement tool 1. Before the factor analysis, it is determined that the scales have a normal distribution and there is no multicollinearity problem across items. Also, there is no missing value in hypothetic data set. As a result of EFA, it is determined that items of measurement tool 1 are gathered under five factors, and they are also under their own factors in parallel with the results of original study. Since an item had high factor loading in more than a factor, it was emitted from the analysis. Factor loading values of the items are between .40-.80. The contributions of items to the total variance are as follows; for first factor 10.63%, for second factor 10.02%, for third factor 8.87%, for fourth factor 8.03%, for fifth factor 6.94% and the total variance explained is 44.49%. In CFA results, which was performed to produce additional evidence for construct validity, the standardized coefficients of items which had a significant t value may change between .32-.70, and the error variance values may change between .50-.90. As a result of the analysis, it is determined that fit indices are S-Bχ^2(366)=699.22, p=.000, χ^2/df=1.91, RMSEA=.037, GFI=.92, NNFI=.96 and SRMR=.049. It is observed that the Cronbach Alfa coefficients which were calculated to determine internal consistency of factor are for the first factor .75, for the second .78, for the third .72, for the fourth .69, for fifth .57. The total Cronbach Alfa coefficient of the tool is .84

Measurement tool 2 is a tool that consists of three sub-scale to measure an affective trait through using four-rating scoring. In original study, EFA was performed to determine psychometric properties of tool and to obtain construct validity evidence, the discriminant validity was investigated in the direction of the scores collected from two different groups. Item-test correlations were calculated to determine item discrimination, test-retest method

was applied to obtain reliability for stability and Cronbach Alfa coefficients were calculated to obtain reliability for internal consistency.

In this study, the analysis of measurement tool 2 through the hypothetical data set starts with EFA. Before the factor analysis, it is determined that the scales have a normal distribution and there is no multicollinearity problem between items. Also, there is no missing value in hypothetic data set. As a result of EFA, it is determined that the items are gathered under three factors. Some items are emitted from the analysis because they give low factor loading value ($\lambda^2 < .32$) or they are overlapped items. The factor loading of items ranges between .45-.75. The contributions of items to the total variance are as follows, for first factor 21.04%, for second 17.98%, for third 8.98%, and total 48%. In CFA results, which was performed to produce an additional evidence for the construct validity, the standardized coefficients of items which have significant t value may change between .45-.74 and their error variance may change between .45-.80. As a result of the analysis, it is determined that fit indices are S-Bχ^2(227)=423.46, p=.000, χ^2/df=1.87, RMSEA=.050, GFI=.89, NNFI=.97 and SRMR=.052. It is seen that Cronbach Alfa coefficients, which were calculated to determine internal consistency of factor are for the first factor .89, for the second factor .84, and for the third factor .65. It is not necessary to calculate the total point within the frame of theoretical and logical view, so the whole scale was not calculated by Cronbach Alfa coefficient.

2.3. Data Analysis

To find answers to the research questions of study, EFA, CFA, Cronbach Alfa analysis, covariance matrices equality test and multi-group CFA were performed. The factor analysis aims to find a few but significant new (common) unrelated variable by combining the variables related with each other in p-variable situation (p-dimensional space). In other words, the factor analysis is a method in which common components are determined and construct dependence is dispelled (Diekhoff, 1992; Gorsuch, 1974; Tatlıdil, 1992; Thompson, 2004; Tucker and MacCallum, 1997). Factor analysis is a technique, which is used to confirm whether the items of a certain scale or sub-scale are gathered under a certain construct or factor (Gable and Wolf, 2001). Beyond reducing variable and naming the emerging factors, the EFA reveals whether the analysis results are similar to the structure of the theory (unobserved latent variables) that enables to figure out the behaviour. After the analysis, a query is made for determining whether the indicators, which are gathered under a certain factor, are indicators of theoretical construct. In CFA, it is firstly aimed to test and confirm the structural hypotheses regarding the relationships between variables. Within this frame, it is focused on examining the relationships between factors and variables, and the relationships between factors in this research through the hypothesis developed. Therefore, the researcher should have the information about the construct of variables that s/he defined in model before the analysis. By this way, the model can be based on a strong theoretical or empirical basis (Raykov and Marcoulides, 2008; Stevens, 1996). Multi-group CFA, which is a special application of CFA can test the measurement and equality of construct models for multi-groups (Brown, 2006). The factor loads of measurement tool consist of measurement properties related to the variables that include constants and error variances. The multi-group CFA makes comparison between two or more groups simultaneously possible by using covariance matrices that are calculated for each compared groups. Thus, measurement invariance or equivalence can be tested by putting equality constraints the parameters of groups (Harrington, 2009).

For the model comparisons in the studies in which the measurement invariance is tested through the multi-group CFA, the first approach of two common approaches is the

comparison between the structural model developed by the hypothesis that there is no meaningful difference between the factor structures for each compared group and the alternative models. In the second approach, the fit between the more constrained nested model and the least constrained comparison model is evaluated by following a stepwise process. Although researcher suggests that evaluation of difference between models should be made between nested models, the comparisons were made for each methods to increase sample numbers and Δ's were evaluated in this study.

Additionally, cut of points for factor loading in EFA are accepted as $\lambda^2 \geq .32$; χ^2 level of acceptance in hypothesis test for significance as .05; since n>300 in each data set the cut of points for multi-group CFA in measurement invariance run for three-model comparison as $\Delta CFI \leq -.01$; as $\Delta SRMR \leq .03$ for weak factorial invariance and as $\Delta SRMR \leq .01$ for strict factorial invariance.

LISREL sample syntax for covariance matrices is in appendix 1, LISREL sample syntax for four models, which are based on for measurement invariance is in appendix 2.

3. Findings

The five-factor structure of measurement tool 1 was tested to determine measurement invariance with multi-group DFA for the groups in the majors of science, health and social science. Before giving the findings of measurement invariance, test statistics, normality tests and reliability coefficients are given in Table 3 in terms of basis assumption of analysis.

Table 3. *Test statistics, normality tests and reliability coefficients of science, health and social science groups*

Major	Factor	n	Mean	Median	Mode	s	Range	Skewness	Kurtosis	α[1]
Science	1	215	17.52	18	24	4.68	18	-.429	-.686	.74
	2	215	16.89	17	15	3.94	17	-.184	-.521	.82
	3	215	21.28	22	23	4.01	19	-.490	-.224	.75
	4	215	16.19	16	20	3.06	13	-.672	-.060	.72
	5	215	15.42	16	17	3.31	15	-.796	.388	.65
	Scale	215	87.29	88	90	12.43	66	-.521	.374	.85
Health	1	212	17.07	18	20	4.80	18	-.422	-.681	.76
	2	212	17.65	18	19	4.03	18	-.433	-.224	.79
	3	212	20.97	21	23	4.42	20	-.637	.135	.73
	4	212	16.09	17	20	3.07	13	-.605	-.129	.68
	5	212	15.14	16	17	3.41	15	-.544	-.202	.62
	Scale	215	86.91	89	97	13.26	66	-.536	.014	.85
Social	1	239	17.79	18	24	4.49	18	-.374	-.766	.75
	2	239	17.37	18	16	4.03	17	-.188	-.676	.76
	3	239	21.89	22	21	4.17	17	-.406	-.674	.74
	4	239	16.27	17	20	3.14	15	-.805	.507	72
	5	239	14.57	14	13	3.36	15	-.313	-.225	.56
	Scale	239	87.90	88	86	13.05	63	-.303	.014	.85

[1] Cronbach Alfa internal consistency coefficient

As can be seen in Table 3, measures of central tendency are relatively close to each other for the groups in the majors of science, health and social science in the level of both sub-scale scores and total scale scores. The fact that skewness and kurtosis coefficients are in the range of ∓ 1 indicate that the distribution is close to normal (Rosenthal and Rosnow, 2008). Although the coefficients are between ∓ 1, it can be said that all of the sub-scales and total scale score points are partly negatively skewed distribution. Accordingly, multi-group CFA, which was performed to determine whether measurement invariance was provided or not for all groups, was computed through asymptotic covariance matrix and $S-B\chi^2$ statistics was

used as base for model fit. On the other hand, it is seen that internal consistency coefficients of science, health and social science groups, which were calculated based on sub-scale and scale scores, are generally in an acceptable level. According to Nunnaly and Bernstein (1994), the reliability coefficient may be accepted for the research if the value is between .70-.80. In all groups, .70 condition is fulfilled with the factors 1-2-3, and 4 at a level of scale. However, this acceptance cannot be provided at the level of factor 5 in all groups. It can be thought that the internal consistency coefficient of sub-scale is low because the number of items is low. The equality of covariance matrices in science, health and social science groups was tested before multi-group CFA.

As a result of the analysis, index values of fit between covariance matrices related to the groups are shown in Table 4.

Table 4. *The equality of covariance matrices of science, health and social science groups*

Groups	S-Bχ^2(df)	p	χ^2/df	RMSEA	GFI	CFI	SRMR
Science, Health and Social	1025.2(870)	.000	1.178	.028 (.020-.035)	.91	.98	.060

As can be seen in Table 4, S-Bχ^2 and degree of freedom are below 2, RMSEA is below .05, GFI is above .090, CFI is above .95 and SRMR is below .08. In this situation it can be said that there is a fit between three covariance matrices.

Multi-group CFA findings for five-factor structure equality of the measurement tool 1 are given in Table 4 for science, health and social science groups.

Table 5. *Findings of multi-group confirmatory factor analysis for science, health and social science groups (maximum possibility)*

	S-Bχ^2(df)[1]	MC[2]	$\Delta\chi^2$(Δdf)	χ^2/df	$\Delta\chi^2$/Δdf	CFI	ΔCFI	SRMR	ΔSRMR	Decision
Science	522.35(366)	–	–	1.427	–	.95	–	.073	–	–
Health	477.91(366)	–	–	1.306	–	.97	–	.064	–	–
Social	536.72(366)	–	–	1.466	–	.95	–	.068	–	–
Model 1[A]	1735.47(1234)	–	–	1.406	–	.95	–	.079	–	–
Model 2[B]	1671.52(1176)	M1–M2	63.95(58)	1.421	-.015	.95	0	.075	.004	H$_0$Accept
Model 3[C]	1859.75(1205)	M1–M3	-124.28(29)	1.543	-.137	.94	.01	.081	-.002	H$_0$Accept
		M2–M3	-188.23(-29)	–	-.122	–	.01	–	-.006	H$_0$Accept
Model 4[D]	1920.04(1265)	M1–M4	-184.57(-31)	1.518	-.112	.94	.01	.085	-.006	H$_0$Accept
		M3–M4	-60.29(-60)	–	**.025**	–	0	–	-.004	H$_0$Accept

[1] p<.05
[2] Model comparison (M=Model)
[A] Configural Invariance (Factor loads, factor correlation and error variance are constant)
[B] Weak Factorial Invariance (Factor loads, factor correlation and error variance are constant)
[C] Strong Factor Invariance (Factor loads and error variance are free, factor correlation is constant)
[D] Strong Factorial Invariance (Error variance is free, factor loads and factor correlation are constant)

Firstly, when the fit indices obtained as a result of CFA which was performed separately for science, health and social science groups are examined, it can be said that fit indices obtained from each of the three groups largely meet the acceptance levels. Accordingly, it can be seen that S-Bχ^2 and the degree of freedom are below 2, CFI is equal to .95 or above this value and SRMR is below .08. After analysing the fit indices in general, it can be said that for five-factor structure of the measurement tool 1 was confirmed separately for science, health and social science groups.

The configural model, which was developed with the hypothesis about there is no significant difference between factor loads, factor correlation and error variance for science, health and social science groups was tested to evaluate measurement invariance. The analysis results show that S-Bχ^2 and the degree of freedom are below 2, CFI is equal to .95 or above this value and SRMR is below .08. Also, after analysing the fit indices in general, it can be acceptable that fit indices of configural model meet the acceptance levels.

When configural invariance (Model 1) and weak factorial invariance (Model 2) models are compared, it is seen that fit gets worse in terms of the ratio of S-B$\Delta\chi^2$ and Δdf. In addition, it can be said that there is no change in ΔCFI value and the change is not significant (n<.025) in ΔSRMR.

When configural invariance (Model 1) and strong factorial invariance (Model 3) models are compared in terms of the ratios of S-B$\Delta\chi^2$ and Δdf, it's seen that the fit gets worse. Besides it can be said that the fit between models gets worse regarding the ΔCFI and ΔSRMR values. On the other hand, when weak factorial invariance (Model 2) and strong factorial invariance (Model 3) models are compared in terms of the second approach, between the ratios of both S-B$\Delta\chi^2$ and Δdf, ΔCFI and ΔSRMR values, it can be stated that the fit gets worse.

Finally, when configural invariance (Model 1) and strict factorial invariance (Model 4) models are compared, it can be seen that the fit gets worse in terms of S-B$\Delta\chi^2$ and Δdf ratios for both models. Besides it can be stated that the fit between models also gets worse in terms of ΔCFI and ΔSRMR values. On the other hand, according to the second approach when strong factorial invariance (Model 3) and strict factorial invariance (Model 4) models are compared, the fit in terms of S-B$\Delta\chi^2$ and Δdf ratios, gets better. In this direction T_s value calculated for difference ratio of S-Bχ^2 is 57.3 and it is confirmed that this value is smaller than the critical value in χ^2 distribution table, $\chi^2_{diff}(60)=79.08$, p>.05. Therefore, it can be said that there is no significant difference between strong factorial invariance and strict factorial invariance models. In other respects, it can be stated that there is no change in ΔCFI value and ΔSRMR value in the direction of the fit gets worse.

In the light of findings outlined above, among the four models, the model that works best based upon covariance matrices in the majors of science, health and social science is configural invariance model developed in assumption of the equality of factor structures. In this context, it is accepted that the five-factor structure of the measurement tool 1 is equal for relevant groups, in other words, measurement invariance is supported.

Measurement invariance for the three-factor structure of measurement tool was tested through multi-group CFA for both female and male groups. Before giving findings about measurement invariance, first in line with the basic assumption of the analysis, test statistics related to relevant groups, test of normality and reliability coefficients were given in Table 6.

Table 6. *Test statistics, tests of normality and reliability coefficients for female and male groups*

Gender	Factor	n	Mean	Median	Mode	S	Range	Skewness	Kurtosis	α[1]
Female	1	220	9.57	8	0	7.45	28	0.51	-0.84	.89
	2	220	14.58	15	12	6.31	26	-0.05	-0.78	.83
	3	220	4.30	4	3	3.14	12	0.55	-0.53	.68
Male	1	133	8.99	7	3	7.39	30	0.82	-0.23	.87
	2	133	13.84	12	9	7.13	27	0.29	-1.04	.84
	3	133	4.55	4	3	3.01	12	0.44	-0.44	.61

[1] Cronbach Alfa internal consistency coefficient

As seen in Table 6, it can be stated that measures of central tendency for female and male groups in the level of both sub-scale and scale total points is close. It is stated that except one distribution of coefficient of skewness and kurtosis, all other distributions even though it is between ∓ 1, points, to some extent, are negative-skewed. Also it is seen in data set for male students that sub-scale points are out of ∓ 1; when kurtosis coefficient is calculated to kurtosis' standard error, the obtained value is still out of ∓ 1.96. In this direction, multi-group CFA, which intends to find out whether the CFA and measurement invariance are confirmed

for each group, is performed over asymptotic covariance matrix and model fit was based on S-Bχ^2. On the other hand, it is seen that internal consistency coefficient calculated on female and male groups' sub-scale and scale points is generally in an acceptable level. In all groups, .70 condition meets with 1st and 2nd factors in the scale level but in both groups, this acceptance cannot be met in 3rd factor level. It can be concluded that the internal consistency coefficient of the relevant scale is low because the number of items (4 items) in the scale is low.

The equality of covariance matrices for female and male groups was tested before the multi-group CFA. In the result of analysis made, index values regarding the fit between covariance matrices for these groups are presented in Table 7.

Table 7. *Equality of covariance for female and male groups*

Groups	S-Bχ^2(df)	p	χ^2/df	RMSEA	GFI	CFI	SRMR
Female & Male	374.04(276)	.000	1.355	.042(.029-.053)	.89	.99	.083

As can be seen in Table 7, the ratio of S-Bχ^2 to the degree of freedom is below 2, RMSEA is below .05, GFI is below .90, CFI is above .95 and SRMR is above .08. It can be stated that when fit indices are assessed in general and GFI and SRMR indices are taken into account, the fit between two variances is moderate.

The multi-group CFA findings related to the equality of the three-factor structure of the measurement tool 2 for female and male groups are shown in Table 8.

Table 8. *Multiple-group confirmatory factor analysis findings of female and male groups (maximum possibility)*

	S-Bχ^2(df)[1]	MC[2]	$\Delta\chi^2$(Δdf)	χ^2/df	$\Delta\chi^2$/Δdf	CFI	ΔCFI	SRMR	ΔSRMR	Decision
Female	349.31(227)	–	–	1.539	–	.98	–	.061	–	
Male	298.51(227)	–	–	1.315	–	.98	–	.068	–	
Model 1	732.94(503)	–	–	1.457	–	.97	–	.100	–	
Model 2	694.38(480)	M1–M2	38.56(23)	1.447	**.010**	.97	0	.083	.017	H$_0$Accept
Model 3	651.04(457)	M1–M3	81.90(46)	1.425	**.032**	.99	**-.02**	.076	**.024**	**H$_0$Reject**
		M2–M3	43.34(23)	–	**.022**	–	**-.02**	–	**.007**	**H$_0$Reject**
Model 4	691.77(480)	M1–M4	41.17(23)	1.441	**.016**	.98	-.01	.098	.002	H$_0$Accept
		M3–M4	-40.73(-23)	–	-.016	–	.01	–	-.022	H$_0$Accept

[1] p<.05
[2] Model Comparison (M=Model)

Firstly, it can be said when fit indices obtained as a result of CFA's performed separately for female and male groups are analysed, fit indices of both groups generally meet the level of acceptance. According to this, it is seen that S-Bχ^2 and degree of freedom ratios are below 2, CFI is above .95 and SRMR is below .08. It can be stated that when fit indices are assessed in general, the three-factor structure of the measurement tool is confirmed separately for female and male groups.

To evaluate the measurement invariance, factor loads, factor correlations, and error variances for female and male groups of the three-factor structure are initially tested with the configural invariance model based on the hypothesis asserting that there is no significant difference among the variables stated. In the result of the analysis made the ratio of S-Bχ^2 to the degree of freedom is below 2, CFI is above .95 but SRMR is above .08. When fit indices are assessed in general, fit indices for the configural invariance model meet the level of acceptance in general.

When configural invariance (Model 1) and weak factorial invariance (Model 2) models are compared, it is seen that the fit gets better in terms of S-B$\Delta\chi^2$ and Δdf ratio for both models. In this direction T$_s$ value calculated for difference ratio of S-Bχ^2 is 41.02 and this value is bigger than the critical value in χ^2 distribution table, χ^2_{diff}(23)=35.17, p<.05. In other words,

there is a significant difference between configural invariance and weak factorial invariance models. On the other hand, it is seen that there is no difference in ΔCFI value, and the change in ΔSRMR is not significant (<.025). When the findings are evaluated in general, based on the results asserting that the difference is not significant in two out of three fit indices, it is decided that fit indices of the configural invariance and the weak factorial invariance do not differ from each other.

When configural invariance (Model 1) and strong factorial invariance (Model 3) models are compared, it is seen that the fit gets better in terms of S-B$\Delta\chi^2$ and Δdf ratio for these two models. In this direction, T_s value calculated for difference ratio of S-Bχ^2 is 78.14 and it has been determined that this value is bigger than the critical value on the χ^2 distribution table, $\chi^2_{diff}(46)=62.83$, p<.05. Hence, there is a significant difference between configural invariance and strong factorial invariance models. Also, when Model 1 and 3 are compared, it can be said that ΔCFI value (<-.01) and ΔSRMR value (>.01) considerably change. On the other hand, according to the second approach, when weak factorial invariance (Model 2) and strong factorial invariance models are compared, the similar results are obtained with the first approach comparison. According to this, the fit between S-B$\Delta\chi^2$ and Δdf percentages gets better.. In this direction T_s value calculated for difference ratio of S-Bχ^2 is 38.37 and this value is bigger than the critical value on the χ^2 distribution table, $\chi^2_{diff}(23)=35.17$, p<.05. In other words, there is a significant difference between weak factorial invariance and strong factorial invariance models. Also in the context of Model 2 and 3, ΔCFI value (<-.01) and ΔSRMR value (>.01) significantly differ.

Finally, when configural invariance (Model 1) and strict factorial invariance (Model 4) models are compared, it is observed that the fit gets better in terms of S-B$\Delta\chi^2$ and Δdf ratio for these two models. In this direction, T_s value calculated for difference ratio of S-Bχ^2 is 37.5 and this value is bigger than the critical value on the χ^2 distribution table, $\chi^2_{diff}(23)=35.17$, p<.05. According to that, there is a significant difference between weak factorial invariance and strict factorial invariance models. But on the other hand it is seen that the change is not significant in ΔCFI (=-.01) and ΔSRMR values (<.01). When the findings are evaluated, since there is no significant difference in two of three fit indices, it was decided that fit indices of the strict factorial invariance model and the configural invariance model don't differ from one another. According to the second approach, when strong factorial invariance (Model 3) and strict factorial invariance (Model 4) models are compared, it can be said that the fit between two models gets worse in terms of S-B$\Delta\chi^2$ and Δdf ratio and ΔCFI and ΔSRMR values.

The findings above reveal that the best working model is the strong factorial invariance model among these four models. Accordingly, it has been accepted that the three-factor structure of the measurement tool 2 is not equal for female and male groups, and measurement invariance cannot be supported.

In this step, different exploratory factor analyses have been computed for the data set obtained from female and male groups. As a result of analysis carried out for the female group, it has been observed that the three factors structure of measurement tool 2 is valid for this group, there is no considerable difference in factor loads (.38 and .75) and the contribution (%49.56) of these factors to the total variance explained. On the other hand, the analysis results, which were performed for male group have revealed that the items have been gathered under four factors. As a result of analysis repeated for these four factors, it is revealed that three items belonging to the second factor is showed up as another factor. In the data set for male group, it has been seen that factor loads (.47 and .83) and the contribution of

the factors to the total variance (57.55%) goes up, however the increase in the total variance explained has occurred because of the rise in the number of factors. When the results of analysis are examined in general, the factor numbers for measurement tool 2 are three for the female group and four for the male group.

4. Discussion and Conclusion

Validity is a concept that is referred to inferences from trait(s) that it measures, but not behalf of the measurement tool's name. Not only the ones who develop and adapt measurement tool but also the researchers who use the tool with a different purpose or different sample from its initial purpose have certain responsibilities to reveal scientific evidences about the validity. However, it may not be sufficient for determining the psychometric properties of the measurement tool through conventional methods in all circumstances. For a particular measurement tool and group, it is a problematic issue for construct validity whether the factor structures that were confirmed empirically by the factor analysis, have the same meaning for independent sub-groups in the sample. Therefore, it might be required for the researchers to test validity for different groups in the sample. It is thereby a crucial psychometric problem to test the equality for specific groups obtained factor structure design by explanatory and/or confirmatory factor analyses in scale development or adaptation studies.

The researchers usually make comparisons across groups to create theoretical information or contribute to existent theoretical knowledge one and naturally want their decisions as correct as possible about the population that they wish to generalize according to their findings. The researchers should examine whether the factor structure of the tool is equal across groups because these comparisons are usually made by scores of the measurement tool. In contrast, in case of inequality of the factor structures defined in the measurement tool across groups, the group scores of these structures do not mean the same. When the factor structures are equal among groups, factor design will be the same for groups and thus it can be evaluated that the collected group scores from scale or sub-scale are valid. That leads to develop a new additional empirical evidence for construct validity of the measurement tool. On the other hand, if psychometric properties of the measurement tools are not equal for groups, the factor structures will vary from one sub-group to another, the comparisons made with scores of measurement tool and the decisions about group will be faulty. Also if there is no empirical evidence, the contribution of the research results to the theory will be doubtful. In this respect, it is beneficial to develop a different perspective on the construct validity, in the framework of this basic problem, in scale development or adaptation studies in the majors of behavioural and educational sciences. Hence, the researcher has the responsibility to test the equality of test scores from factor structure of the tool and sub-scale for the compared groups. Therefore, the evaluations related to empirical evidence about the validity of measurement tool are never considered as "last word". Validating a tool requires everlasting effort.

This study aimed at developing a sample for what sort of decisions could be made for two different factor designs that either of them supports measurement invariance. In this respect, it was determined that the five-factor structure of the measurement tool 1, regarded as a first sample, supported measurement invariance for the undergraduate students in the majors of science, health, and social sciences. Based on EFA and CFA results, it can be claimed that the five-factor structure of the tool has high validity for the entire of the sample. With the multi-group CFA, the relevant factor design was approved to be separately valid for students of science, medical and social sciences. This result recommends that this factor design is valid for the whole sample and all of the science majors (any of these groups) like science,

health, and social sciences separately, so that the measurement tool 1 has high construct validity for the groups that maintain their academic lives in different majors.

As a second sample, the three-factor structure of the measurement tool 2 supports measurement invariance for high school students within the context of gender. In such cases, the researchers should consider the relevant factor structure as unequal for groups, therefore, they need to take into account the different factor designs or possibility of bias for each comparison. Therefore, the researchers are suggested to run the EFA for each group. In fact, it has been revealed that three factor structure is valid for females but not in males, in which the items are gathered under four factor in this study. It is required to give cues to make the findings more meaningful (reason of hiding names of scales are explained in method section). In the EFA that is run for males, the three items of the second factor named as stress sub-scale have been observed to be gathered under a new factor. When these items are examined, three of them are about "negative reaction showed in blocking situation". These items, which are symptoms of stress for females, are loaded under a new factor called "intolerantness to blocking" for males. This finding reveals that the structure for female and male groups are different, in other words, doesn't have the same psychological meaning for these groups. The researchers can produce different forms of measurement tool in this situation (e.g. female form-male form) and suggest these forms for the ones who study in the sub-groups of the sample. Although this situation causes a problem for practicality, it plays a crucial role in the construct validity.

The researchers can test whether the measurement tool, based on the theoretical basis of the trait which it intends to measure is equal for more than a group or a sample like age, gender, socio-economic level, class, education level, academic major, subcultures in a society, international comparisons, experimental researchers, and different occupational groups. It is surely beyond doubt that the evidences about whether the measurement invariance is supported for different groups will strengthen psychometric properties of the tool, and will therefore increase the validity of the results presented in the direction of the research findings. As a result, the contribution to the production of theoretical knowledge or existing theoretical knowledge accumulation will enhance as well.

References

Baumgartner, H. & Steenkamp, J. E. M. (1998). Multi–group latent variable models for varying numbers of items and factors with cross–national and longitudinal applications. *Marketing Letters*, *9*(1), 21-35.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. (First Edition). New York: Guilford Publications, Inc.

Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamada Eğitim Yönetimi*, *32*, 470-483.

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2012). *Bilimsel araştırma yöntemleri* (11. Baskı). Ankara: Pegem Akademi.

Büyüköztürk, Ş. (2014). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* (On Dokuzuncu Baskı). Ankara: Pegem Yayıncılık.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (Second Edition). New York: Taylor & Francis Group.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, *20*(4), 872-882.

Byrne, B. M. (2006). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. (Second Edition). California: Sage Publications, Inc.

Byrne, B. M. & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 287-321.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464-504.

Cheung, G., & Rensvold, R. (2000). *Testing measurement invariance using critical values of fit indices: A Monte Carlo study*. Retrieved May 20, 2004, from http://www.aom.pace.edu/rmd/cheung_files/ cheung.htm

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness–of–fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. (First Edition). Orlando: Holt, Rinehart and Winston, Inc.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Diekhoff, G. (1992). *Statistics for the social and behavioral sciences: Univariate, bivariate, and multivariate* (First Edition). Dubuque, IA: William C. Brown Publishers.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43, 121-149.

Dunn, G., Everitt, B. & Pickles, A. (1993). *Modeling covariances and latent variables using EQS* (First Edition). London: Chapman & Hall.

Everitt, B. S. & Howell, D. C. (2005). *Encyclopedia of statistics in behavioral sciences*. Chichester: John Wiley & Sons, Ltd.

Floyd, F. J. & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286-299.

Gable, R. K. & Wolf, M. B. (2001). *Instrument development in the affective domain: Measuring attitudes, and values in corporate and scholl settings* (Second Edition). London: Kluwer Academic Publishers.

Green, S. B., Salkind, N. J. & Akey, T. M. (1997). *Using SPSS for windows: Analyzing and understanding data*. New Jersey: Prentice Hall, Inc.

Gregorich, S. (2006). Do self report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11-3), 78-94.

Gorsuch, R. L. (1974). *Factor analysis*. (First Edition). Philadelphia: W. B. Saunders Company.

Harrington, D. (2009). *Confirmatory factor analysis* (First Edition). New York: Oxford University Press, Inc.

Jonson, J. L. & Plake, B. S. (1998). A historical comparison of validity standards and validity practices. *Educational and Psychological Measurement*. *58*(5), 736-754.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409-426.

Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Lincolnwood: Scientific Software International, Inc.

Jöreskog, K. G. & Sörbom, D. (2001). *LISREL 8: User's reference guide*. Lincolnwood: Scientific Software International, Inc.

Jöreskog, K. G., Sörbom, D., Toit, M. & Toit, S. (2000). *LISREL 8: New statistical features*. Lincolnwood: Scientific Software International, Inc.

Kline, P. (2000). *The handbook of psychological testing* (Second Edition). London: Taylor & Francis Group.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (Second Edition). New York: Guilford Publications, Inc.

Lee, S. Y. & Leung, T. K. (1982). Covariance structure analysis in several populations. *Psychometrika*, 47, 297-308.

Marcoulides, G. A. & Schumacker, R. E. (1996). *Advanced structural equation modeling: ıssues and techniques* (First Edition). New Jersey: Lawrence Erlbaum Associates, Inc.

Maruyama, G. M. (1998). *Basics of structural equation modeling* (First Edition). Thousand Oaks, CA: SAGE Publications, Inc.

Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are internet and paper-and-pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods, 10,* 322-345.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525-543.

Nunnaly, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (Third Edition). New York: McGraw-Hill, Inc.

Raykov, T. & Marcoulides, G. A. (2008). An introduction to applied multivariate analysis (First Edition). New York: Taylor & Francis Group.

Rosenthal, R & Rosnow, R. L. (2008). *Essential of behavioral research* (Third Edition). New York: McGraw-Hill, Inc.

Sass, D. A., Schmitt, T. A. & Marsh H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 167-180.

Satorra, A. & Bentler, P. M. (2011). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika: A Journal of Quantitative Psychology*, *66*(4), 507-514.

Schmitt, N. & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*, 210-222.

Schumacker, R. E. & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling* (First Edition). New Jersey: Lawrence Erlbaum Associates, Inc.

Spini, D. (2003). Measurement equivalence of 10 value types from the Schwartz value survey across 21 countries. *Journal of Cross-Cultural Psychology*, *34*(1), 3-23.

Stapleton, C. D. (1997). Basic contepts and procedures of confirmatory factor analysis. Austin: The Annual Meeting of the Southwest Educational Research Association.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*, 893-898.

Stevens, J. (1996). Applied multivariate statistics for social sciences (Third Edition). New Jersey: Lawrence Erlbaum Associates, Inc.

Tatlıdil, H. (1992). *Uygulamalı çok değişkenli istatistiksel analiz* (Birinci Baskı) Ankara: Engin Yayınları.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications* (First Edition). Washington: American Psychological Association.

Toit, M. & Toit, S. (2001). *Interactive LISREL: User's guide*. Lincolnwood: Scientific Software International, Inc.

Tucker, L. R. & MacCallum, R. C. (1997). Exploratory factor analysis. (Online Edition) Web: http://quantrm2.psy.ohio-state.edu/maccallum/factornew.htm adresinden 19 Şubat 2007'de alınmıştır.

Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (pp. 653-771). Needham Heights, MA: Allyn & Bacon.

Urbina, S. (2004). *Essentials of psychological testing* (First Edition). New Jersey: Wiley & Sons, Inc.

Vandenberg, R. J. & Lance, C. E. (2000). A Review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4-70.

Westen, D. & Rosenthal, R. (2005). Improving construct validity: Cronbach, Meehl, and neurath's ship. *Psychological Assessment*, *17*(4), 409-41.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association.

Wu. A. D., Li, Z. & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multigroup confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, *12*(3), 1-26.

Appendix 1: LISREL Syntax Example for Test of Covariance Matrices Equality
```
SCALE I EQUALITY TEST OF COVARIANCE MATRIX (ACADEMIC MAJOR)
Group SCIENCE:
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
Latent Variables: f1-f30
Relationships:
V1 = 1*f1
V2 = 1*f2
V3 = 1*f3
V4 = 1*f4
V5 = 1*f5
V6 = 1*f6
V7 = 1*f7
V8 = 1*f8
V9 = 1*f9
V10 = 1*f10
V11 = 1*f11
V12 = 1*f12
V13 = 1*f13
V14 = 1*f14
V15 = 1*f15
V16 = 1*f16
V17 = 1*f17
V18 = 1*f18
V19 = 1*f19
V20 = 1*f20
V21 = 1*f21
V22 = 1*f22
V23 = 1*f23
V24 = 1*f24
V25 = 1*f25
V26 = 1*f26
V27 = 1*f27
V28 = 1*f28
V29 = 1*f29
V30 = 1*f30
Set the Error Variances of V1-V30 to zero
Group HEALTHCARE:
Observed Variables: V1-V30
Covariance Matrix from File HEALTHCARE.COV
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
Latent Variables: f1-f30
Group HUMANITIES:
Observed Variables: V1-V30
Covariance Matrix from File HUMANITIES.COV
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
Latent Variables: f1-f30
End of Problem
```

Appendix 2: LISREL Syntaxes Example for Test of Measurement Invariance
```
Model A
Group 1: Testing Equality Of Factor Structures
Model A: Factor Loadings, Factor Correlation, Error Variances Invariant
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Asymptotic Covariance Matrix from File SCIENCE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5
```

```
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
End of Problem
Model B
Group 1: Testing Equality Of Factor Structures
Model B: Factor Correlation and Error Variances Invariant
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Asymptotic Covariance Matrix from File SCIENCE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
End of Problem
Model C
Group 1: Testing Equality Of Factor Structures
Model C: Factor Correlation Invariant
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Asymptotic Covariance Matrix from File SCIENCE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
```

```
Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Set the Error Variances of V1-V30 free
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Set the Error Variances of V1-V30 free
End of Problem
Model D
Group 1: Testing Equality Of Factor Structures
Model D: Factor Loadings and Factor Correlation Invariant
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Asymptotic Covariance Matrix from File SCIENCE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
Set the Error Variances of V1-V30 free
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
Set the Error Variances of V1-V30 free
End of Problem
```

MEASUREMENT INVARIANCE: CONCEPT AND IMPLEMENTATION

Güçlü Şekercioğlu [iD]

Akdeniz University, Turkey

guclus@akdeniz.edu.tr

Güçlü Şekercioğlu has had his Ph.D. in Measurement and Evaluation from Ankara University, Turkey. He is a teaching staff at Akdeniz University, Education Faculty. He conducts research in multivariate statistics, measurement invariance, differential item functioning and psychological measurement in culture context.

MEASUREMENT INVARIANCE: CONCEPT AND IMPLEMENTATION

Güçlü Şekercioğlu

guclus@akdeniz.edu.tr

Abstract

An empirical evidence for independent samples of a population regarding measurement invariance implies that factor structure of a measurement tool is equal across these samples; in other words, it measures the intended psychological trait within the same structure. In this case, the evidence of construct validity would be strengthened within the frame of the scores obtained from the tool. When measurement invariance is not supported, the researchers should consider the possibility of the different factor designs for each group. Ignoring such a situation brings forward the probability about differentiation of the trait(s) measured by measurement tool for that/those group(s), so it causes to suspect the validity of the scores obtained from the tool. The aim of this study is to examine measurement invariance in the context of the conceptual foundations of multi-group confirmatory factor analysis, and discuss the subject through the results from two hypothetical data set that one supports measurement invariance, but the other does not. As a result of analysis performed in this direction, it is determined that the five-factor design derived from the first data set is equal across the groups in the majors of science, health, and social science. It is also concluded that the three-factor design obtained from the secondary data set is not equal for female and male groups. Besides, the exploratory factor analysis performed for female and male groups separately shows that the three-factor design of the tool is valid for females, but the number of factors was four in males. When the factor design for male group is examined, it is determined that the three items in the second factor separate significantly. That leads to the conclusion that it is crucial to test measurement invariance in studies regarding the determination of the psychometric properties of the tool.

Keywords: measurement invariance, equality of factor structures, multi-group confirmatory factor analysis, structural equation modeling

1. Introduction

The major problem in behavioural and educational science studies, which aim developing the psychological measurement tools, cultural adaptation of a tool developed in another culture, using the tool for a different purpose or for a different sample, is to demonstrate the validity of the empirical evidence on the psychometric properties of the tool. In this direction the researchers, within the framework of these fundamental problems, are obliged to question whether the tool measures the trait(s) what it intends to measure properly and precisely. Further examination related to psychometric properties of measurement tool and all other analyses based on the scores obtained from the measurement tool (ANOVA, regression, etc.) has been carried out after the validity of the evidence put forward and decision are taken in this direction. According to Nunnally and Bernstein (1994), the validity of each usage must be documented by empirical evidence even though a measurement result may be valid for more than one purpose. Therefore, the test authors and users should not assume that the validity of evidence cannot change (Crocker and Algina, 1986).

One of the most important dimensions of the validity of scores obtained from psychological measurement tools is the construct validity. In the report of testing standards published in 1954 it was discussed that the concept of validity, actually all types of validity should be assembled under the roof of construct validity (Cronbach and Meehl, 1955; Jonson and Plake, 1998; Urbina, 2004; Westen and Rosenthal, 2005). Similarly, Kline (2000) states that the construct validity includes other approaches as well, thus all types of validity are related to the assessment of construct validity. The factor analysis is one of the most commonly used techniques in the studies which aim to determine the psychometric properties of a measurement tool in behavioural and educational science, in order to obtain evidence of construct validity. According to Büyüköztürk (2002, 2014) the factor analysis is a multivariate statistics, which aims to find and explore conceptually meaningful fewer new variables (factor, component) by bringing a large number of inter-related variables together. The factor analysis can be considered under two headings, which are exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) discussed under the concept of structural equation modelling (SEM).

CFA, which is based on testing of theories about the latent variables, and used in advanced research, is a very sophisticated technique (Ullman, 2001). In this analysis, a predefined and constrained construct is tested whether it is confirmed as a model. It is also occasionally used to mean the confirmation of the theoretical structure (Maruyama, 1998). In this context, the determination of the construct validity for CFA is emphasized as a very powerful method (Floyd and Wideman, 1995; Kline, 2005; Stapleton, 1997).

Multi-group confirmatory factor analysis (multi-group CFA) is also a spesific practice area in CFA. This analysis enables to test the equality of structural parameters for more than one group simultaneously. In this context, the assessment of equality between the groups in terms of factor structure is also termed as *measurement invariance*. Additionally, examining the fitness of structure brings about the concept of testing population heterogeneity. It is possible to encounter different terms for different tests of measurement invariance tests including equality test of factor structures, metric and factorial invariance in the literature (Brown, 2006).

Nowadays, interest of the researchers in social sciences towards measurement invariance is gradually increasing. In a plain defination, measurement invariance is the description of whether the structures of measurement tool are equal across individuals from different groups. This concept has a critical importance in comparing groups. When measurement invariance is not supported between the groups, it is not possible to interpret the findings that reveal differences concerning these groups. If the researcher does not have the evidence for measurement invariance, then the existence of different psychometric responses for scale items more than one group cannot be known. Measurement invariance analyses are used in intercultural comparison for groups speaking different languages in a culture, scale adaptation studies, the comparison of groups with different academic achievement, the comparison of employee groups in different areas of industry, comparisons based on gender and are also used to compare a control group and an experimental group in empirical research (Cheung and Rensvold, 2002). The frequently asked question on the use of psychological measurement tools is whether the factor design ensued as a result of factor analysis of the measurement tools valid for groups, which differentiated at such a level that may impact the measurement process concerning the ethnic characteristics, age or the way they respond to the items. In fact, the fundamental issue here is whether the measurement tool measures the same structural properties for different groups or not. When the factor structure is not equal across groups, naturally it is not possible to make meaningful comparisons between groups based on the factor scores. On the contrary, when measurement equivalence is supported

empirically, it is concluded that the group differences are completely reflected in terms of latent traits evaluated by factors. In this concept, the studies, which aim at determining equality of the measurement tools' factor structure, are becoming more and more substantial because cultural, developmental and contextual impacts related to the psychological structural traits have become well-known by the researchers recently (Floyd and Widaman, 1995). In addition to social science studies, using multi-group CFA becomes increasingly common in other majors such as psychology, education, management and organization, marketing, and communication, especially ones which carried out studies based on cross-cultural comparisons.

As Jöreskog, Sörbom and Toit (2000) claim, the factor structure of developed or adapted scale based on fundamental data set obtained from different groups or samples can be tested whether it is equal for more than one group or not concerning the national, territorial, regional, cultural or socio-economic status of the groups. It is highly functional to test the equality of factor structures for a scale or different numbers of items for more than a group. Thus, factors or structural relationships can be tested simultaneously whether they are equal across different samples (Baumgartner and Steenkamp, 1998).

According to Marcoulides and Schumacker (1996) in multi-group CFA, the question of "*is each group measured under the same structure?*" is investigated and this examination is carried out within the framework of the measurement model defined in advance. Similarly, Kline (2005) stated that, the focus of multi-group CFA is to test whether measurement invariance is supported for different groups within the same latent variables. This concept is defined as invariance of the psychometric properties of a scale across groups in the context of modelling in the literature of psychometry.

Determining whether the measurement invariance is supported for different groups or not has a critical role in the development of psychometric properties of psychological measurement tools. That implies whether the items of the same structure and all structures can be used for the sub-groups of a population. Likewise, the subject of testing measurement invariance plays a crucial role in terms of defining the generalizability of psychological structure across groups with differinf variables such as different cultures, age groups and genders. The equality tests of latent means, which are included in the analysis group, shows similarity with the comparison of observed group averages through t-test and ANOVA (Brown, 2006).

According to Byrne (2006), the researchers often seek answers to any of the following five questions for evidence related to the multi-group equality: (i) do certain structures of the items on the measurement tool work equally across different groups? In other words, does the measurement model have a group equality? (ii) is the factor structure of the tool or theoretical structure measured by multiple scales equal for each level of the group? (iii) are the paths of the experimental structures equal across the groups? (iv) does the latent means in the model for a particular structure vary between groups? and (v) is the factor structure of a measurement tool equal for independent samples of the population? The author particularly emphasizes that there could be a cross-validation study in his last question. The analysis results reach the conclusion that if the factor structures are not equal between the groups, the validity of interpretations based on a comparison of scores for these groups decreases.

According to Brown (2006), the process steps below should be followed in the evaluation of the multi-group CFA and measurement invariance: (i) performing CFA for each group included in the analysis separately, (ii) testing the equality of structures simultaneously (factor loadings, factor correlations and error variances constant), (iii) testing the equality of the factor structures (factor loadings free; factor correlations and error variances constant),

(iv) testing the equality of factor structures and the error variances indicators (factor loadings and error variances free; factor correlations constant), (v) testing the equality errors variances of indicator (error variances free, factor loadings and factor correlations constant), (vi) testing the equality of factor variances, (vii) testing the equality of factor covariances (if more than one factor), and (viii) testing the equality of latent means. Hereunder, the first step is one of the multi-group CFA's assumption. The processes between the second and fifth steps are about testing measurement invariance, and the processes between the sixth and eighth steps are about testing the population heterogeneity.

1.1. Measurement Invariance Test and Models

Before computing the multi-group CFA, first of all, correlation or covariance matrix of the groups in the same sample is evaluated by comparing each other. In other words, before setting up the configural invariance model (Model 1), the establishment of the test equality of covariance matrices (Model 0) must be made. If the equality of covariance matrices is provided for each group $(\Sigma^g = \Sigma^{g'})$ the configural invariance model can be developed and tested. The equality of covariance structures of the groups should be discussed only after the null hypothesis (H_0) has been rejected. Subsequently, the models for other hypothesizes should be tested separately. The configural model derived from different groups should be defined in the same sample. Thus, the defined model for each group of multi-group analysis would be simultaneously tested. In this case, it is expected to see high fitness between correlation or covariance matrices of different groups (Brown, 2006; Byrne, 2006; Dunn, Everitt and Pickles, 1993; Vandenberg and Lenca, 2000). In general, the measurement invariance is tested with four basic models. These models are summarized in Table 1 (adapted from Cheung and Rensvold, 2002).

Table 1. *Measurement invariance models*

Models	Hypothesis	Hypothesis Name	Symbolic Statement	Process
1	H_{form}	Configural invariance	$\Lambda^{(1)}_{form} \therefore \Lambda^{(2)}_{form}$	Invariance is supported for all groups regarding construct and items. Factor loadings, factor correlations, and error variances are equal for all groups.
2	H_Λ	Weak Factorial Invariance (Metric Invariance)	$\Lambda^{(1)} \therefore \Lambda^{(2)}$	Invariance is supported for all groups regarding factor correlations and error variances. The factor loadings have been freed for groups.
3	H_λ	Strong Factorial Invariance (Scalar Invariance)	$\lambda^{(1)}_{ij} \therefore \Lambda^{(2)}_{ij}$	Invariance is supported for all groups regarding factor correlations. The factor loadings and error variances have been freed for groups.
4	$H_{\Lambda,\Theta(\delta)}$	Strict Factorial Invariance (Residual Variance Invariance)	$\Theta^{(1)}_\delta \therefore \Theta^{(2)}_\delta$	Invariance is supported for all groups regarding factor loadings and correlations. The error variances have been freed for groups.

1.1.1. Configural invariance (baseline model)

Developing a configural invariance (also known as baseline model) begins with identifying and testing the model, which was developed within the framework of a specific hypothesis for each group. In this context, the number of sub-scales in configural invariance model for each group (e.g. factors), the positions of the items (e.g. which factors the items are loaded) and correlations between sub-scales (e.g. setting such factors covariance) are determined. Secondly, the validity of the configural invariance model is tested separately for each group. Ideally, the model is expected to well fit and significant. However, the evidence, which shows a well fit, provides the information to the researcher that only the factor structure is similar but does not give any certain information about the equality of factors for each group. The evidence act as a design for invariance tests to be carried out subsequently. This model has two important functions: (i) the parameters are tested simultaneously for all groups, (ii) equal initial value is generated for the integration of configural invariance model

for testing (Byrne, 2008). Hence, the criterion, which will be obtained from further models to be tested, is occurred. In this model, invariance, regarding structure and items are supported for all groups (factor loadings, factor correlations and error variances are equal for all groups). When weak, strong or strict factorial invariance hypothesizes are rejected, the "factor structure is equal across all groups" hypothesis, which is developed within the framework of configural invariance, is accepted.

1.1.2. Weak factorial invariance (metric invariance)

In this model, the equality of factor loadings (λ), ($\Lambda^1 = \Lambda^2 = ... = \Lambda^G$) is tested for all groups. (Spini, 2003; Vandenberg and Lence, 2000). If the fit, which is obtained from weak factorial invariance test, is better than the fit of configural invariance, configural invariance model is rejected. In other words, it indicates that the equivalence is not supported. According to Byrne and Stewart (2006) although measurement units are identical for groups in terms of underlying factors (e.g. factor loads), it constitutes one of the constraints of this model because scaling (e.g. intercepts) is not identical. Therefore, Meredith (1993) describes this invariance level as *weak factorial invariance*. This invariance is tested,

$$M_g \cong \tau_g \hat{\tau}'_g + \hat{\Lambda}(\alpha_g \; \acute{\alpha}_g + \hat{\Psi}_g)\hat{\Lambda} + \hat{\Theta}_{\varepsilon g} = \hat{M}'_g$$

with this equation (Widaman and Reise, 1997).

1.1.3. Strong factorial invariance (scalar invariance)

It is tested whether the regression constant (τ) of observed variables on the latent variables is equal across groups ($\tau^1 = \tau^2 = ... = \tau^3$) (Schmitt and Kuljanin, 2008). In this model, there are a series of additional constraints described in weak factorial invariance. These additional constraints include the intercepts of the variables that are observed in the matrices τ_g. If estimations are problematic in terms of invariance on groups, subscript g on matrix τ is removed. In this case, invariance is tested,

$$M_g \cong \hat{\tau}\hat{\tau}' + \hat{\Lambda}(\alpha_g \acute{\alpha}_g + \hat{\Psi}_g)\hat{\Lambda} + \hat{\Theta}_{\varepsilon g} = \hat{M}'_g$$

with this equation (Widaman and Reise, 1997).

1.1.4. Strict factorial invariance (residual variance invariance)

In this last model of the measurement invariance, about error terms across the groups $H_{\Lambda\phi}$ model limits ($H_{\Lambda\phi\theta}$) model equally ($\theta^1 = \theta^2 = \cdots = \theta^G$). With the addition of this constraint, testing the hypothesis of equality of measurement errors becomes possible for independent samples of the population. If the error variances are equal, it means the items have equal reliability in terms of groups (Spini, 2003). Strict factorial invariance is also created through strong factorial invariance as it occurs in strong factorial invariance created through the weak factorial model constraints. These additional constraints are defined as strict factorial invariance, which contains unique factorial invariance in $\hat{\Theta}_{\varepsilon g}$ matrix or measurement errors. This invariance is tested,

$$M_g \cong \hat{\tau}\hat{\tau}' + \hat{\Lambda}(\alpha_g \; \acute{\alpha}_g + \hat{\Psi}_g)\hat{\Lambda} + \hat{\Theta}_\varepsilon = \hat{M}_{g'}$$

with this equation (Widaman and Reise, 1997).

It should be noted that there are varrious classification in the related literature. Therefore, it is worth to consider following aspects suggested by Meredith (1993) and Dimitrov (2010), in the testing process of the equality of factor structures across groups, metric invariance is the general name of weak factorial invariance, strong factorial (scale invariance) and strict factorial invariance (invariance of error variance) models. However, there are some research

in literature that discuss the weak factorial invariance with the term of metric invariance. (Gregorich, 2006; Meade, Michels and Lautenschlager, 2007; Schmitt and Kuljanin, 2008; Spini, 2003; Vandenberg and Lance, 2000; Wu, Li and Zumbo, 2007). Besides, Cheung and Rensvold (2002) used the terms *metric invariance on construct-level* for weak factorial invariance, *item-level metric invariance* for strong factorial invariance and *error variance invariance* for strict factorial invariance.

Multi-group CFA for measurement invariance can be computed with such software statistical programs like LISREL, Amos, SAS/STAT, Mplus and EQS. The analysis starts with the creation of separate covariance matrices for the levels of the groups. It can be carried out by typing the syntax analysis in LISREL program or by following the instructions prescribed by the program (Toit and Toit, 2001). Measurement invariance is carried out in four models. The syntax samples of these models are named as EX10A.SPL, EX10B.SPL, EX10C.SPL, and EX10D.SPL in LISREL program. In the first model (Model 1), also known as configural invariance model, factor loads of structure(s), correlations and error variance are assumed to be equal and the analysis is run in this regard. The configural invariance model, which is a fundamental model for the equality of factor structure, is developed with the hypothesis that factor structures are equal (H_0=*There is no difference between factor structures*). In order to make comparisons with model defined in the analysis, a second alternative model named as weak factorial invariance model (Model 2) is analysed. In the weak factor invariance model, freeing the factor loads for each group, keeping the factor correlations and error variances constant are discussed. In the third alternative model strong factorial invariance (Model 3) factor loads and error variances for each level of the group are released, factor correlations are kept constant. The last and fourth model of measurement invariance is strict factor invariance model (Model 4). In this model while error variances are released, factor loads and factor correlations are kept constant (Byrne, 2010; Jöreskog and Sörbom, 1993; Toit and Toit, 2001).

1.2. Model Comparisons in the Decision of Measurement Invariance

In multi-group CFA invariance test, constrained and unconstrained model are compared. In terms of availability of different values for each group in constrained model, model parameters (e.g. factor loads) are not constrained in this model. The parameters have the same value for all groups in constrained model. When the fit of unconstrained model is better than the constrained one, it implies that constrained model is incorrect. In other words, if the unconstrained model fits better when the constrained parameters are released, they are allowed to get different values for each group, and the constrained model developed within the invariance hypothesis framework is rejected (Cheung and Rensvold, 2000).

For comparisons of models with multi-group CFA in the studies in which measurement invariance is tested, it can be said that there are two widely used approaches in literature. The first one is the comparison between configural invariance model developed with the hypothesis that there is no difference in factor structure for each group and alternative models (e.g. weak factorial invariance, strong factorial invariance and strict factorial invariance models). Hereunder, the first comparison is made between configural invariance model and weak factorial invariance model (model 1 and 2), the second is between configural invariance model and strong factorial invariance model (model 1 and 3), and the last one is between configural invariance model and strict factorial invariance model (model 1 and 4). According to this approach, in the case of equality of fit between any alternative model and configural invariance or in the event of deterioration, the configural invariance model developed with the hypothesis that there is no difference in factor structure for each group is accepted. On the other hand, if the alternative model indexes differ from the configural invariance indexes

significantly (in favour of alternative models), H_0 hypothesis is rejected. In this case, the equality of factor structure and thus, the measurement invariance cannot be supported (Byrne, 2010; Jöreskog and Sörbom, 1993; Toit and Toit, 2001). In the second approach, the comparisons are performed by following stepwise process. According to this, the analysis starts with less limited models and then the models are assessed by using nested χ^2 method (Brown, 2006). Accordingly, in comparison to nested models; $H_{form} > H_{\Lambda} > H_{\lambda} > H_{\Lambda,\Theta(\delta)}$ is used as base. In other words, comparisons are made between configural invariance model and weak factorial model (model 1 and 2), weak factorial invariance model and strong factorial invariance model (model 2 and 3), strong factorial invariance and strict factorial invariance (model 3 and 4). According to Van de Vijyer and Leung (1997) if the fit of nested models is equal, more constrained model is frequently accepted. If this is not the case, the equality hypothesis is rejected (as cited in Spini, 2003). Cheung and Rensvold (2002) also suggest another comparison containing only one difference from the first approach. Although the first two comparison is the same, the authors suggest a comparison between weak factorial invariance and strict factorial invariance (model 2 and 4).

1.3. Decision Making of Measurement Invariance

While deciding whether the factor structures are equal for each group, the significance level of χ^2 matrix is required and the level should above .05 value, in other words, a non-significance value p is expected. This situation means that the covariance matrix of each of the defined groups do not differ significantly, thereby measurement invariance is supported. According to Jöreskog and Sörbom (1993), examples of acceptability of fit indices provided in Table 2 might be used for decision.

Table 2. *Acceptance of equality of factor structure in multi-group confirmatory factor analysis*

Problem	χ^2	df	p value	Decision
A	38.08	10	0.000	Reject
B	1.52	2	0.468	Accept
C	8.77	4	0.067	Accept
D	21.55	8	0.006	Reject
E	38.22	11	0.000	Reject

As seen in Table 2, models A, D, and E in which significance value p is a problem, are rejected whereas problem B and C are accepted. The criteria determined in the developing first years of multi-group CFA have been questioned over time. χ^2 has a possibility to increase its significance value if the number of samples increases, therefore, alternative models are investigated whether to accept the fit of factor structures within the model framework or not to assess the fit between covariance matrices. Among these, firstly, the value of χ^2 and degree of freedom should be compared. In this regard the χ^2 value obtained from the more constrained model, χ^2 value from less constrained model and the "delta" value (delta means the difference and its symbols is Δ) which is between the degree of freedom are calculated. $\Delta\chi^2$ and Δdf values are determined with this calculation. The significance level of χ^2 value obtained from this determination, is controlled in the level of p<.01 or p<.05 by comparing the critical values in the distribution table of χ^2 (Byrne, 2010; Jöreskog, 1971; Kline, 2005; Lee and Leung, 1982; Steiger, 2007; Van den Bergh and Van Ranst, 1998). In this case, H_0 and H_1 hypotheses can be developed in the following format:

H_0: There is no significant difference between the more constrained model and less constrained models in terms of fit.

H₁: There is a significant difference between the more constrained model and less constrained models in terms of fit.

In this respect, if $\Delta\chi^2$, which is calculated on the basis of χ^2 differences in a particular Δdf level, is less than critical table values, H_0 is accepted. In other words, there is no significant difference between two models in terms of fit, therefore, the researcher can make a decision about measurement invariance based on $\Delta\chi^2$. On the other hand, if $\Delta\chi^2$, which is calculated on the basis of χ^2 differences in a particular Δdf level, is more than critical table values, H_0 is rejected. Thus, there is a significant difference between two models in terms of fit, and if this difference is in favour of the alternative hypothesis, the researcher can assume that measurement invariance is not provided on the basis of $\Delta\chi^2$.

In many studies in which analysis of SEM concept is applied, the distribution(s) may be remote from normal within certain tolerances. In the absence of normality in large samples, χ^2 value (S-Bχ^2) obtained from Satorra-Bentler correction produces close values to the χ^2 that is produced when the number of people in the sample and the distribution of the produces is normal. S-Bχ^2 is a rather reliable statistical test used to evaluate covariance structure models in various distributions and sample sizes (Byrne, 2006; Everitt and Howell, 2005). As in the other SEM analyses, such as multi-group CFA, which is carried out to obtain evidence of measurement invariance, S-Bχ^2 can only be calculated if the distribution of each group is far from the normal distribution. In multi-group CFA, which is carried out with the maximum likelihood method, T_s value should be calculated for S-Bχ^2 scaled difference in terms of evidence of measurement invariance between nested models. T_s is calculated

$$T_s = (T_0 - T_1) / c_d$$

with this equation. T_0 is the normal maximum likelihood χ^2 value for nested model, T_1 is the normal maximum likelihood χ^2 value for comparison (less constrained model) model, and c_d is the degree of difference test correction. c_d is calculated

$$c_d = [(d_0 * c_0) - (d_1 * c_1)] / (d_0 - d_1)$$

with this equation. d_0 is the degree of freedom of nested model, d_1 is the degree of freedom of comparison model, c_0 is the correction degree of nested model, and c_1 is the correction degree of comparison model. c_0 and c_1 are calculated

$$c_0 = T_0 / T_0^* \text{ and } c_1 = T_1 / T_1^*$$

with this equation. T_0^* is S-Bχ^2 value of nested model, on the other hand T_1^* is S-Bχ^2 value of comparison model. By comparing T_s, which is calculated for S-Bχ^2 difference degrees, with the critical values in χ^2 distribution table, it can be determined whether measurement invariance is supported (Brown, 2006; Satorra and Bentler, 2011).

Recently, it is widely used as an alternative to utilize from fit indices as well as to evaluate the χ^2 differences among the models in many research due to a large number of n. According to Cheung and Rensvold (2002) it is inadvisable to reject null hypothesis in case of obtaining an insignificant χ^2 value. χ^2 is statistically sensitive test for large samples, however, it is not a practical test for model fit. In such case, alternative fit indices should be offered for χ^2. The comparative fit indices (CFI, NNFI / TLI, RMSEA etc.) are among the most frequently recommended ones. Within this framework, it is observed that many goodness of fit indices are commonly used together to evaluate general fit of the model and to report it. ΔGFI's are used as an alternative for χ^2 in multi-group CFA which is performed to determine whether the factor structures are equal or not. As in χ^2, the configural model whose factor loads, factor correlation and error variance are released in covariance matrices of groups, in other

words, the model which is developed with the hypothesis that factor structures are equal, is the basic model like in alternative fit indices. For the evidence of measurement invariance, the differences between models can be evaluated with the comparison of indices such as ΔRMSEA, ΔCFI, ΔGamma Hat, ΔMc, ΔIFI, ΔAIC, ΔEVCI, ΔNFI, ΔTLI, and ΔSRMR.

The fit values are expected to become better for the equality of factor structures when the parameters like factor loads and error variance in covariance matrices of the group are released together or one by one. With this regard, the differences are evaluated by comparing the indices (e.g. ΔSRMR, ΔCFI and ΔRMSEA) between configural model and other alternative models or nested models. The configural model set up with the hypothesis that there is no significant difference between factor structures of each group is accepted if the fit indices of alternative models are lower than the ones in configural model. On the other hand, if the fit indices of other alternative model are higher than the ones in configural model or nested model, the fit across models is evaluated whether it differs significantly or not.

Cheung and Rensvold (2000; 2002) suggested cut-off points for ΔCFI significance level between modes in terms of measurement invariance after carrying out a study by using Monte Carlo method. Hereunder, when ΔCFI≥−.01 is provided, then configural invariance model is accepted. In contrast to this situation, if ΔCFI is between –.01 and –.02, there will be increasing doubt about invariance. If it is more than -.02 it can be said that the difference between constrained and unconstrained model will increase. In this situation, configural model is rejected. In this context, it is decided that the factor structures are not equal and therefore an alternative model should be sought. In addition, the critical values of ΔGamma hat and ΔMcDonald NFI are –.001 and –.02.

Chen (2007) suggested cut-off points for decision of measurement invariance by considering situations like sample size of CFI, RMSEA and SRMR indices and sample sizes in groups after carrying out a study, which aimed at testing sensitivity of goodness of fit indices through Monte Carlo method. Accordingly, it can be concluded that measurement invariance cannot be supported (case of noninvariance) if sample size is small (n<300), sample sizes of groups are not equal, pattern of variance is the same, there is a relationship like ΔCFI≤−.005, ΔRMSEA≥.010 or ΔSRMR≥.025 between groups in terms of weak factorial invariance test, and there is a relationship like ΔCFI≥-.005, ΔRMSEA≥.010 or ΔSRMR≥.005 between groups in terms of strong factorial invariance or strict factorial invariance. On the other hand, measurement invariance can be supported when sample size is sufficient (n>300), numbers of groups compared are equal, there is a relationship like ΔCFI≥−.010, ΔRMSEA≥.015 or ΔSRMR≥.030 between groups in terms of weak factorial invariance test, and there is a relationship like ΔCFI≥−.010, ΔRMSEA≥.015 between groups in terms of strong factorial invariance or strict factorial invariance.

An important point to be considered in assessing multi-group CFA comparison of the four basic models is type I and type II error possibilities. If the sample is small for a null hypothesis, type I error is likely occurred. However, if the sample is getting larger for alternative hypothesis, the difference of fit will be extended. In that case, type II error is likely occurred. For this reason, to minimize the type I and type II error possibility, the cut-off points should be determined efficiently (Hu and Bentler, 1998). In their maximum possibility $\Delta\chi^2$ studies which were performed with the indicators acting as continuous variables, French and Finch (2008) controlled type I error in the level of .01 and .05 between different models and sample numbers. The researchers revealed that the power of $\Delta\chi^2$ has a positive correlation with sample size, indicator number of each factors and factor number. Meade and Bauer (2007) also extrapolated the same results about $\Delta\chi^2$ (as cited in Sass,

Schmitt and Marsh, 2014). There is no doubt that this case is valid for other delta fit indices as well. However, this study didn't include detailed discussions on that subject because it was beyond the scope.

1.4. Objectives

Researchers of behavioural and educational sciences provide evidence through a sample on the validity of scores obtained from developed or adapted psychological measurement tools. After revealing the psychometric properties of the measurement tools, measurement process can be practiced on an independent group in the same sample and various decisions may be taken by means of obtained scores in the same or a different study. The fact that a measurement tool with confirmed factor structure for a sample may not be valid for the independent sub-groups in the relevant sample is a probability that researchers should pay attention. In such a case, the validity of decisions to be taken with scores obtained from groups will be suspicions. Within this scope, this research aims to discuss the conceptual basis of multi-group CFA in measurement invariance in terms of basic concepts and to introduce the subject through two hypothetic data set that one supports measurement invariance, but the other doesn't, for the researchers aiming to determine the psychometric properties of a measurement tool. Thereby, a new perspective will be introduced to the researchers aiming to determine the psychometric properties of measurement tools, suggestions about decisions to be taken for the tool without equalized factor structure will be asserted. Accordingly, the present study searches answers to the following research questions:

1. Is the five-factor structure of measurement tool 1 equal across the groups of science, health and social sciences?
2. Is the three-factor structure of measurement tool 2 equal across groups of males and females?

This research is limited to measurement invariance (measurement of configural invariance, weak factorial invariance, strong factorial invariance and strict factorial invariance). The heterogeneity of the population (factor variance invariance, factor covariance invariance and latent means invariance) is not included in the research.

2. Method

This study examines the method of multi-group CFA for the evidence concerning measurement invariance through two data set consisting of equal and unequal factor structure. Considering the findings of the study, the current study has the characteristics of correlational research concerning equality of factor structure for independent groups in two samples and due to the discussions on generation of construct validity evidences. The correlational studies analyse the relationship between two or more variables without intervening in these variables under any circumstances. These studies are the ones that are effective on revealing the relationships and determining the levels of relationships between variables and provides necessary cues for conducting high-level research on these relationships (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz and Demirel, 2012).

2.1. Research Data

The ready-made data was used in this study. They consist of two data set (equal and unequal factor structured) that the researchers collected them from his previous researches. The first hypothetic data set that measurement invariance is supported consist of 666 undergraduate students. When the distribution of the participants is examined based on scientific major, 32.28% (n=215) science, 31.83% (n=212) of health and 35.89% (n=239) of social science. The other hypothesis data set that measurement invariance is not supported

consist of 353 high school students. The distribution in terms of participants' gender is as follows, 62.32% (n=220) female, 37.68% (n=133) male.

2.2. Data Collection Tools

The study consists of two hypothetic data set, which are the subjects of measurement invariance analyses and the scores obtained from two different measurement tools. Some items were emitted from the tool in line with the results of EFA and CFA that were run on the data set collected from the participants. Moreover, the factor design differed for male participants in the second data collection tool whose factor design was not equal. The main purpose of this study is not to determine or discuss the psychometric properties of aforementioned tools. However the present study focuses on presenting the multi-group CFA in terms of measurement invariance through two hypothetic data set in which measurement invariance both was supported and was not, and creating a new view of validity for the researchers who aim to measure the psychometric properties of a measurement tool. Therefore, it is not appropriate to give the names of the tools and sub-scales in view of the probability that because they can form basis for further studies. For this reason, the data collection tools were mentioned as measurement tool 1 and measurement tool 2, and limited information about the psychometric properties of the tools was given because it was not wanted to reveal the tool.

Measurement tool 1 is a tool that consists of five sub-scales to measure an effective trait through using four point rating. In the original study, EFA and CFA were performed to determine psychometric properties of the tool in terms of gathering evidence about construct validity, concurrent validity was examined by comparing with a criterion score, to obtain reliability evidence for stability a test-retest method was run, and lastly to obtain reliability evidence for internal consistency, Cronbach alfa coefficients were calculated. In conclusion, it can be said that the scores obtained from measurement invariance tool 1 have a high level of validity.

This study starts with EFA to obtain construct validity evidence through hypothetic data set of measurement tool 1. Before the factor analysis, it is determined that the scales have a normal distribution and there is no multicollinearity problem across items. Also, there is no missing value in hypothetic data set. As a result of EFA, it is determined that items of measurement tool 1 are gathered under five factors, and they are also under their own factors in parallel with the results of original study. Since an item had high factor loading in more than a factor, it was emitted from the analysis. Factor loading values of the items are between .40-.80. The contributions of items to the total variance are as follows; for first factor 10.63%, for second factor 10.02%, for third factor 8.87%, for fourth factor 8.03%, for fifth factor 6.94% and the total variance explained is 44.49%. In CFA results, which was performed to produce additional evidence for construct validity, the standardized coefficients of items which had a significant t value may change between .32-.70, and the error variance values may change between .50-.90. As a result of the analysis, it is determined that fit indices are S-Bχ^2(366)=699.22, p=.000, χ^2/df=1.91, RMSEA=.037, GFI=.92, NNFI=.96 and SRMR=.049. It is observed that the Cronbach Alfa coefficients which were calculated to determine internal consistency of factor are for the first factor .75, for the second .78, for the third .72, for the fourth .69, for fifth .57. The total Cronbach Alfa coefficient of the tool is .84

Measurement tool 2 is a tool that consists of three sub-scale to measure an affective trait through using four-rating scoring. In original study, EFA was performed to determine psychometric properties of tool and to obtain construct validity evidence, the discriminant validity was investigated in the direction of the scores collected from two different groups. Item-test correlations were calculated to determine item discrimination, test-retest method

was applied to obtain reliability for stability and Cronbach Alfa coefficients were calculated to obtain reliability for internal consistency.

In this study, the analysis of measurement tool 2 through the hypothetical data set starts with EFA. Before the factor analysis, it is determined that the scales have a normal distribution and there is no multicollinearity problem between items. Also, there is no missing value in hypothetic data set. As a result of EFA, it is determined that the items are gathered under three factors. Some items are emitted from the analysis because they give low factor loading value ($\lambda^2 < .32$) or they are overlapped items. The factor loading of items ranges between .45-.75. The contributions of items to the total variance are as follows, for first factor 21.04%, for second 17.98%, for third 8.98%, and total 48%. In CFA results, which was performed to produce an additional evidence for the construct validity, the standardized coefficients of items which have significant t value may change between .45-.74 and their error variance may change between .45-.80. As a result of the analysis, it is determined that fit indices are S-Bχ^2(227)=423.46, p=.000, χ^2/df=1.87, RMSEA=.050, GFI=.89, NNFI=.97 and SRMR=.052. It is seen that Cronbach Alfa coefficients, which were calculated to determine internal consistency of factor are for the first factor .89, for the second factor .84, and for the third factor .65. It is not necessary to calculate the total point within the frame of theoretical and logical view, so the whole scale was not calculated by Cronbach Alfa coefficient.

2.3. Data Analysis

To find answers to the research questions of study, EFA, CFA, Cronbach Alfa analysis, covariance matrices equality test and multi-group CFA were performed. The factor analysis aims to find a few but significant new (common) unrelated variable by combining the variables related with each other in p-variable situation (p-dimensional space). In other words, the factor analysis is a method in which common components are determined and construct dependence is dispelled (Diekhoff, 1992; Gorsuch, 1974; Tatlıdil, 1992; Thompson, 2004; Tucker and MacCallum, 1997). Factor analysis is a technique, which is used to confirm whether the items of a certain scale or sub-scale are gathered under a certain construct or factor (Gable and Wolf, 2001). Beyond reducing variable and naming the emerging factors, the EFA reveals whether the analysis results are similar to the structure of the theory (unobserved latent variables) that enables to figure out the behaviour. After the analysis, a query is made for determining whether the indicators, which are gathered under a certain factor, are indicators of theoretical construct. In CFA, it is firstly aimed to test and confirm the structural hypotheses regarding the relationships between variables. Within this frame, it is focused on examining the relationships between factors and variables, and the relationships between factors in this research through the hypothesis developed. Therefore, the researcher should have the information about the construct of variables that s/he defined in model before the analysis. By this way, the model can be based on a strong theoretical or empirical basis (Raykov and Marcoulides, 2008; Stevens, 1996). Multi-group CFA, which is a special application of CFA can test the measurement and equality of construct models for multi-groups (Brown, 2006). The factor loads of measurement tool consist of measurement properties related to the variables that include constants and error variances. The multi-group CFA makes comparison between two or more groups simultaneously possible by using covariance matrices that are calculated for each compared groups. Thus, measurement invariance or equivalence can be tested by putting equality constraints the parameters of groups (Harrington, 2009).

For the model comparisons in the studies in which the measurement invariance is tested through the multi-group CFA, the first approach of two common approaches is the

comparison between the structural model developed by the hypothesis that there is no meaningful difference between the factor structures for each compared group and the alternative models. In the second approach, the fit between the more constrained nested model and the least constrained comparison model is evaluated by following a stepwise process. Although researcher suggests that evaluation of difference between models should be made between nested models, the comparisons were made for each methods to increase sample numbers and Δ's were evaluated in this study.

Additionally, cut of points for factor loading in EFA are accepted as $\lambda^2 \geq .32$; χ^2 level of acceptance in hypothesis test for significance as .05; since n>300 in each data set the cut of points for multi-group CFA in measurement invariance run for three-model comparison as ΔCFI$\leq$-.01; as ΔSRMR$\leq$.03 for weak factorial invariance and as ΔSRMR$\leq$.01 for strict factorial invariance.

LISREL sample syntax for covariance matrices is in appendix 1, LISREL sample syntax for four models, which are based on for measurement invariance is in appendix 2.

3. Findings

The five-factor structure of measurement tool 1 was tested to determine measurement invariance with multi-group DFA for the groups in the majors of science, health and social science. Before giving the findings of measurement invariance, test statistics, normality tests and reliability coefficients are given in Table 3 in terms of basis assumption of analysis.

Table 3. *Test statistics, normality tests and reliability coefficients of science, health and social science groups*

Major	Factor	n	Mean	Median	Mode	s	Range	Skewness	Kurtosis	α^1
Science	1	215	17.52	18	24	4.68	18	-.429	-.686	.74
	2	215	16.89	17	15	3.94	17	-.184	-.521	.82
	3	215	21.28	22	23	4.01	19	-.490	-.224	.75
	4	215	16.19	16	20	3.06	13	-.672	-.060	.72
	5	215	15.42	16	17	3.31	15	-.796	.388	.65
	Scale	215	87.29	88	90	12.43	66	-.521	.374	.85
Health	1	212	17.07	18	20	4.80	18	-.422	-.681	.76
	2	212	17.65	18	19	4.03	18	-.433	-.224	.79
	3	212	20.97	21	23	4.42	20	-.637	.135	.73
	4	212	16.09	17	20	3.07	13	-.605	-.129	.68
	5	212	15.14	16	17	3.41	15	-.544	-.202	.62
	Scale	215	86.91	89	97	13.26	66	-.536	.014	.85
Social	1	239	17.79	18	24	4.49	18	-.374	-.766	.75
	2	239	17.37	18	16	4.03	17	-.188	-.676	.76
	3	239	21.89	22	21	4.17	17	-.406	-.674	.74
	4	239	16.27	17	20	3.14	15	-.805	.507	72
	5	239	14.57	14	13	3.36	15	-.313	-.225	.56
	Scale	239	87.90	88	86	13.05	63	-.303	.014	.85

[1] Cronbach Alfa internal consistency coefficient

As can be seen in Table 3, measures of central tendency are relatively close to each other for the groups in the majors of science, health and social science in the level of both sub-scale scores and total scale scores. The fact that skewness and kurtosis coefficients are in the range of ∓ 1 indicate that the distribution is close to normal (Rosenthal and Rosnow, 2008). Although the coefficients are between ∓ 1, it can be said that all of the sub-scales and total scale score points are partly negatively skewed distribution. Accordingly, multi-group CFA, which was performed to determine whether measurement invariance was provided or not for all groups, was computed through asymptotic covariance matrix and S-Bχ^2 statistics was

used as base for model fit. On the other hand, it is seen that internal consistency coefficients of science, health and social science groups, which were calculated based on sub-scale and scale scores, are generally in an acceptable level. According to Nunnaly and Bernstein (1994), the reliability coefficient may be accepted for the research if the value is between .70-.80. In all groups, .70 condition is fulfilled with the factors 1-2-3, and 4 at a level of scale. However, this acceptance cannot be provided at the level of factor 5 in all groups. It can be thought that the internal consistency coefficient of sub-scale is low because the number of items is low. The equality of covariance matrices in science, health and social science groups was tested before multi-group CFA.

As a result of the analysis, index values of fit between covariance matrices related to the groups are shown in Table 4.

Table 4. *The equality of covariance matrices of science, health and social science groups*

Groups	S-Bχ^2(df)	p	χ^2/df	RMSEA	GFI	CFI	SRMR
Science, Health and Social	1025.2(870)	.000	1.178	.028 (.020-.035)	.91	.98	.060

As can be seen in Table 4, S-Bχ^2 and degree of freedom are below 2, RMSEA is below .05, GFI is above .090, CFI is above .95 and SRMR is below .08. In this situation it can be said that there is a fit between three covariance matrices.

Multi-group CFA findings for five-factor structure equality of the measurement tool 1 are given in Table 4 for science, health and social science groups.

Table 5. *Findings of multi-group confirmatory factor analysis for science, health and social science groups (maximum possibility)*

	S-Bχ^2(df)[1]	MC[2]	$\Delta\chi^2$(Δdf)	χ^2/df	$\Delta\chi^2$/Δdf	CFI	ΔCFI	SRMR	ΔSRMR	Decision
Science	522.35(366)	–	–	1.427	–	.95	–	.073	–	–
Health	477.91(366)	–	–	1.306	–	.97	–	.064	–	–
Social	536.72(366)	–	–	1.466	–	.95	–	.068	–	–
Model 1[A]	1735.47(1234)	–	–	1.406	–	.95	–	.079	–	–
Model 2[B]	1671.52(1176)	M1–M2	63.95(58)	1.421	-.015	.95	0	.075	.004	H_0Accept
Model 3[C]	1859.75(1205)	M1–M3	-124.28(29)	1.543	-.137	.94	.01	.081	-.002	H_0Accept
		M2–M3	-188.23(-29)	–	-.122	–	.01	–	-.006	H_0Accept
Model 4[D]	1920.04(1265)	M1–M4	-184.57(-31)	1.518	-.112	.94	.01	.085	-.006	H_0Accept
		M3–M4	-60.29(-60)	–	**.025**	–	0	–	-.004	H_0Accept

[1] p<.05
[2] Model comparison (M=Model)
[A] Configural Invariance (Factor loads, factor correlation and error variance are constant)
[B] Weak Factorial Invariance (Factor loads, factor correlation and error variance are constant)
[C] Strong Factor Invariance (Factor loads and error variance are free, factor correlation is constant)
[D] Strong Factorial Invariance (Error variance is free, factor loads and factor correlation are constant)

Firstly, when the fit indices obtained as a result of CFA which was performed separately for science, health and social science groups are examined, it can be said that fit indices obtained from each of the three groups largely meet the acceptance levels. Accordingly, it can be seen that S-Bχ^2 and the degree of freedom are below 2, CFI is equal to .95 or above this value and SRMR is below .08. After analysing the fit indices in general, it can be said that for five-factor structure of the measurement tool 1 was confirmed separately for science, health and social science groups.

The configural model, which was developed with the hypothesis about there is no significant difference between factor loads, factor correlation and error variance for science, health and social science groups was tested to evaluate measurement invariance. The analysis results show that S-Bχ^2 and the degree of freedom are below 2, CFI is equal to .95 or above this value and SRMR is below .08. Also, after analysing the fit indices in general, it can be acceptable that fit indices of configural model meet the acceptance levels.

When configural invariance (Model 1) and weak factorial invariance (Model 2) models are compared, it is seen that fit gets worse in terms of the ratio of S-B$\Delta\chi^2$ and Δdf. In addition, it can be said that there is no change in ΔCFI value and the change is not significant (n<.025) in ΔSRMR.

When configural invariance (Model 1) and strong factorial invariance (Model 3) models are compared in terms of the ratios of S-B$\Delta\chi^2$ and Δdf, it's seen that the fit gets worse. Besides it can be said that the fit between models gets worse regarding the ΔCFI and ΔSRMR values. On the other hand, when weak factorial invariance (Model 2) and strong factorial invariance (Model 3) models are compared in terms of the second approach, between the ratios of both S-B$\Delta\chi^2$ and Δdf, ΔCFI and ΔSRMR values, it can be stated that the fit gets worse.

Finally, when configural invariance (Model 1) and strict factorial invariance (Model 4) models are compared, it can be seen that the fit gets worse in terms of S-B$\Delta\chi^2$ and Δdf ratios for both models. Besides it can be stated that the fit between models also gets worse in terms of ΔCFI and ΔSRMR values. On the other hand, according to the second approach when strong factorial invariance (Model 3) and strict factorial invariance (Model 4) models are compared, the fit in terms of S-B$\Delta\chi^2$ and Δdf ratios, gets better. In this direction T_s value calculated for difference ratio of S-Bχ^2 is 57.3 and it is confirmed that this value is smaller than the critical value in χ^2 distribution table, $\chi^2_{diff}(60)=79.08$, p>.05. Therefore, it can be said that there is no significant difference between strong factorial invariance and strict factorial invariance models. In other respects, it can be stated that there is no change in ΔCFI value and ΔSRMR value in the direction of the fit gets worse.

In the light of findings outlined above, among the four models, the model that works best based upon covariance matrices in the majors of science, health and social science is configural invariance model developed in assumption of the equality of factor structures. In this context, it is accepted that the five-factor structure of the measurement tool 1 is equal for relevant groups, in other words, measurement invariance is supported.

Measurement invariance for the three-factor structure of measurement tool was tested through multi-group CFA for both female and male groups. Before giving findings about measurement invariance, first in line with the basic assumption of the analysis, test statistics related to relevant groups, test of normality and reliability coefficients were given in Table 6.

Table 6. *Test statistics, tests of normality and reliability coefficients for female and male groups*

Gender	Factor	n	Mean	Median	Mode	S	Range	Skewness	Kurtosis	α[1]
Female	1	220	9.57	8	0	7.45	28	0.51	-0.84	.89
	2	220	14.58	15	12	6.31	26	-0.05	-0.78	.83
	3	220	4.30	4	3	3.14	12	0.55	-0.53	.68
Male	1	133	8.99	7	3	7.39	30	0.82	-0.23	.87
	2	133	13.84	12	9	7.13	27	0.29	-1.04	.84
	3	133	4.55	4	3	3.01	12	0.44	-0.44	.61

[1] Cronbach Alfa internal consistency coefficient

As seen in Table 6, it can be stated that measures of central tendency for female and male groups in the level of both sub-scale and scale total points is close. It is stated that except one distribution of coefficient of skewness and kurtosis, all other distributions even though it is between ∓1, points, to some extent, are negative-skewed. Also it is seen in data set for male students that sub-scale points are out of ∓1; when kurtosis coefficient is calculated to kurtosis' standard error, the obtained value is still out of ∓1.96. In this direction, multi-group CFA, which intends to find out whether the CFA and measurement invariance are confirmed

for each group, is performed over asymptotic covariance matrix and model fit was based on S-Bχ^2. On the other hand, it is seen that internal consistency coefficient calculated on female and male groups' sub-scale and scale points is generally in an acceptable level. In all groups, .70 condition meets with 1st and 2nd factors in the scale level but in both groups, this acceptance cannot be met in 3rd factor level. It can be concluded that the internal consistency coefficient of the relevant scale is low because the number of items (4 items) in the scale is low.

The equality of covariance matrices for female and male groups was tested before the multi-group CFA. In the result of analysis made, index values regarding the fit between covariance matrices for these groups are presented in Table 7.

Table 7. *Equality of covariance for female and male groups*

Groups	S-Bχ^2(df)	p	χ^2/df	RMSEA	GFI	CFI	SRMR
Female & Male	374.04(276)	.000	1.355	.042(.029-.053)	.89	.99	.083

As can be seen in Table 7, the ratio of S-Bχ^2 to the degree of freedom is below 2, RMSEA is below .05, GFI is below .90, CFI is above .95 and SRMR is above .08. It can be stated that when fit indices are assessed in general and GFI and SRMR indices are taken into account, the fit between two variances is moderate.

The multi-group CFA findings related to the equality of the three-factor structure of the measurement tool 2 for female and male groups are shown in Table 8.

Table 8. *Multiple-group confirmatory factor analysis findings of female and male groups (maximum possibility)*

	S-Bχ^2(df)[1]	MC[2]	$\Delta\chi^2$(Δdf)	χ^2/df	$\Delta\chi^2$/Δdf	CFI	ΔCFI	SRMR	ΔSRMR	Decision
Female	349.31(227)	–	–	1.539	–	.98	–	.061	–	
Male	298.51(227)	–	–	1.315	–	.98	–	.068	–	
Model 1	732.94(503)	–	–	1.457	–	.97	–	.100	–	
Model 2	694.38(480)	M1–M2	38.56(23)	1.447	**.010**	.97	0	.083	.017	H$_0$Accept
Model 3	651.04(457)	M1–M3	81.90(46)	1.425	**.032**	.99	**-.02**	.076	**.024**	**H$_0$Reject**
		M2–M3	43.34(23)	–	**.022**	–	**-.02**	–	**.007**	**H$_0$Reject**
Model 4	691.77(480)	M1–M4	41.17(23)	1.441	**.016**	.98	-.01	.098	.002	H$_0$Accept
		M3–M4	-40.73(-23)	–	-.016	–	.01	–	-.022	H$_0$Accept

[1] p<.05
[2] Model Comparison (M=Model)

Firstly, it can be said when fit indices obtained as a result of CFA's performed separately for female and male groups are analysed, fit indices of both groups generally meet the level of acceptance. According to this, it is seen that S-Bχ^2 and degree of freedom ratios are below 2, CFI is above .95 and SRMR is below .08. It can be stated that when fit indices are assessed in general, the three-factor structure of the measurement tool is confirmed separately for female and male groups.

To evaluate the measurement invariance, factor loads, factor correlations, and error variances for female and male groups of the three-factor structure are initially tested with the configural invariance model based on the hypothesis asserting that there is no significant difference among the variables stated. In the result of the analysis made the ratio of S-Bχ^2 to the degree of freedom is below 2, CFI is above .95 but SRMR is above .08. When fit indices are assessed in general, fit indices for the configural invariance model meet the level of acceptance in general.

When configural invariance (Model 1) and weak factorial invariance (Model 2) models are compared, it is seen that the fit gets better in terms of S-B$\Delta\chi^2$ and Δdf ratio for both models. In this direction T$_s$ value calculated for difference ratio of S-Bχ^2 is 41.02 and this value is bigger than the critical value in χ^2 distribution table, χ^2_{diff}(23)=35.17, p<.05. In other words,

there is a significant difference between configural invariance and weak factorial invariance models. On the other hand, it is seen that there is no difference in ΔCFI value, and the change in ΔSRMR is not significant (<.025). When the findings are evaluated in general, based on the results asserting that the difference is not significant in two out of three fit indices, it is decided that fit indices of the configural invariance and the weak factorial invariance do not differ from each other.

When configural invariance (Model 1) and strong factorial invariance (Model 3) models are compared, it is seen that the fit gets better in terms of S-BΔχ^2 and Δdf ratio for these two models. In this direction, T_s value calculated for difference ratio of S-Bχ^2 is 78.14 and it has been determined that this value is bigger than the critical value on the χ^2 distribution table, $\chi^2_{diff}(46)=62.83$, p<.05. Hence, there is a significant difference between configural invariance and strong factorial invariance models. Also, when Model 1 and 3 are compared, it can be said that ΔCFI value (<-.01) and ΔSRMR value (>.01) considerably change. On the other hand, according to the second approach, when weak factorial invariance (Model 2) and strong factorial invariance models are compared, the similar results are obtained with the first approach comparison. According to this, the fit between S-BΔχ^2 and Δdf percentages gets better.. In this direction T_s value calculated for difference ratio of S-Bχ^2 is 38.37 and this value is bigger than the critical value on the χ^2 distribution table, $\chi^2_{diff}(23)=35.17$, p<.05. In other words, there is a significant difference between weak factorial invariance and strong factorial invariance models. Also in the context of Model 2 and 3, ΔCFI value (<-.01) and ΔSRMR value (>.01) significantly differ.

Finally, when configural invariance (Model 1) and strict factorial invariance (Model 4) models are compared, it is observed that the fit gets better in terms of S-BΔχ^2 and Δdf ratio for these two models. In this direction, T_s value calculated for difference ratio of S-Bχ^2 is 37.5 and this value is bigger than the critical value on the χ^2 distribution table, $\chi^2_{diff}(23)=35.17$, p<.05. According to that, there is a significant difference between weak factorial invariance and strict factorial invariance models. But on the other hand it is seen that the change is not significant in ΔCFI (=-.01) and ΔSRMR values (<.01). When the findings are evaluated, since there is no significant difference in two of three fit indices, it was decided that fit indices of the strict factorial invariance model and the configural invariance model don't differ from one another. According to the second approach, when strong factorial invariance (Model 3) and strict factorial invariance (Model 4) models are compared, it can be said that the fit between two models gets worse in terms of S-BΔχ^2 and Δdf ratio and ΔCFI and ΔSRMR values.

The findings above reveal that the best working model is the strong factorial invariance model among these four models. Accordingly, it has been accepted that the three-factor structure of the measurement tool 2 is not equal for female and male groups, and measurement invariance cannot be supported.

In this step, different exploratory factor analyses have been computed for the data set obtained from female and male groups. As a result of analysis carried out for the female group, it has been observed that the three factors structure of measurement tool 2 is valid for this group, there is no considerable difference in factor loads (.38 and .75) and the contribution (%49.56) of these factors to the total variance explained. On the other hand, the analysis results, which were performed for male group have revealed that the items have been gathered under four factors. As a result of analysis repeated for these four factors, it is revealed that three items belonging to the second factor is showed up as another factor. In the data set for male group, it has been seen that factor loads (.47 and .83) and the contribution of

the factors to the total variance (57.55%) goes up, however the increase in the total variance explained has occurred because of the rise in the number of factors. When the results of analysis are examined in general, the factor numbers for measurement tool 2 are three for the female group and four for the male group.

4. Discussion and Conclusion

Validity is a concept that is referred to inferences from trait(s) that it measures, but not behalf of the measurement tool's name. Not only the ones who develop and adapt measurement tool but also the researchers who use the tool with a different purpose or different sample from its initial purpose have certain responsibilities to reveal scientific evidences about the validity. However, it may not be sufficient for determining the psychometric properties of the measurement tool through conventional methods in all circumstances. For a particular measurement tool and group, it is a problematic issue for construct validity whether the factor structures that were confirmed empirically by the factor analysis, have the same meaning for independent sub-groups in the sample. Therefore, it might be required for the researchers to test validity for different groups in the sample. It is thereby a crucial psychometric problem to test the equality for specific groups obtained factor structure design by explanatory and/or confirmatory factor analyses in scale development or adaptation studies.

The researchers usually make comparisons across groups to create theoretical information or contribute to existent theoretical knowledge one and naturally want their decisions as correct as possible about the population that they wish to generalize according to their findings. The researchers should examine whether the factor structure of the tool is equal across groups because these comparisons are usually made by scores of the measurement tool. In contrast, in case of inequality of the factor structures defined in the measurement tool across groups, the group scores of these structures do not mean the same. When the factor structures are equal among groups, factor design will be the same for groups and thus it can be evaluated that the collected group scores from scale or sub-scale are valid. That leads to develop a new additional empirical evidence for construct validity of the measurement tool. On the other hand, if psychometric properties of the measurement tools are not equal for groups, the factor structures will vary from one sub-group to another, the comparisons made with scores of measurement tool and the decisions about group will be faulty. Also if there is no empirical evidence, the contribution of the research results to the theory will be doubtful. In this respect, it is beneficial to develop a different perspective on the construct validity, in the framework of this basic problem, in scale development or adaptation studies in the majors of behavioural and educational sciences. Hence, the researcher has the responsibility to test the equality of test scores from factor structure of the tool and sub-scale for the compared groups. Therefore, the evaluations related to empirical evidence about the validity of measurement tool are never considered as "last word". Validating a tool requires everlasting effort.

This study aimed at developing a sample for what sort of decisions could be made for two different factor designs that either of them supports measurement invariance. In this respect, it was determined that the five-factor structure of the measurement tool 1, regarded as a first sample, supported measurement invariance for the undergraduate students in the majors of science, health, and social sciences. Based on EFA and CFA results, it can be claimed that the five-factor structure of the tool has high validity for the entire of the sample. With the multi-group CFA, the relevant factor design was approved to be separately valid for students of science, medical and social sciences. This result recommends that this factor design is valid for the whole sample and all of the science majors (any of these groups) like science,

health, and social sciences separately, so that the measurement tool 1 has high construct validity for the groups that maintain their academic lives in different majors.

As a second sample, the three-factor structure of the measurement tool 2 supports measurement invariance for high school students within the context of gender. In such cases, the researchers should consider the relevant factor structure as unequal for groups, therefore, they need to take into account the different factor designs or possibility of bias for each comparison. Therefore, the researchers are suggested to run the EFA for each group. In fact, it has been revealed that three factor structure is valid for females but not in males, in which the items are gathered under four factor in this study. It is required to give cues to make the findings more meaningful (reason of hiding names of scales are explained in method section). In the EFA that is run for males, the three items of the second factor named as stress sub-scale have been observed to be gathered under a new factor. When these items are examined, three of them are about "negative reaction showed in blocking situation". These items, which are symptoms of stress for females, are loaded under a new factor called "intolerantness to blocking" for males. This finding reveals that the structure for female and male groups are different, in other words, doesn't have the same psychological meaning for these groups. The researchers can produce different forms of measurement tool in this situation (e.g. female form-male form) and suggest these forms for the ones who study in the sub-groups of the sample. Although this situation causes a problem for practicality, it plays a crucial role in the construct validity.

The researchers can test whether the measurement tool, based on the theoretical basis of the trait which it intends to measure is equal for more than a group or a sample like age, gender, socio-economic level, class, education level, academic major, subcultures in a society, international comparisons, experimental researchers, and different occupational groups. It is surely beyond doubt that the evidences about whether the measurement invariance is supported for different groups will strengthen psychometric properties of the tool, and will therefore increase the validity of the results presented in the direction of the research findings. As a result, the contribution to the production of theoretical knowledge or existing theoretical knowledge accumulation will enhance as well.

References

Baumgartner, H. & Steenkamp, J. E. M. (1998). Multi–group latent variable models for varying numbers of items and factors with cross–national and longitudinal applications. *Marketing Letters*, *9*(1), 21-35.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. (First Edition). New York: Guilford Publications, Inc.

Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamada Eğitim Yönetimi*, *32*, 470-483.

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş. ve Demirel, F. (2012). *Bilimsel araştırma yöntemleri* (11. Baskı). Ankara: Pegem Akademi.

Büyüköztürk, Ş. (2014). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* (On Dokuzuncu Baskı). Ankara: Pegem Yayıncılık.

Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (Second Edition). New York: Taylor & Francis Group.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, *20*(4), 872-882.

Byrne, B. M. (2006). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. (Second Edition). California: Sage Publications, Inc.

Byrne, B. M. & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 287-321.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464-504.

Cheung, G., & Rensvold, R. (2000). *Testing measurement invariance using critical values of fit indices: A Monte Carlo study*. Retrieved May 20, 2004, from http://www.aom.pace.edu/rmd/cheung_files/ cheung.htm

Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness–of–fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233-255.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. (First Edition). Orlando: Holt, Rinehart and Winston, Inc.

Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*, 281-302.

Diekhoff, G. (1992). *Statistics for the social and behavioral sciences: Univariate, bivariate, and multivariate* (First Edition). Dubuque, IA: William C. Brown Publishers.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43, 121-149.

Dunn, G., Everitt, B. & Pickles, A. (1993). *Modeling covariances and latent variables using EQS* (First Edition). London: Chapman & Hall.

Everitt, B. S. & Howell, D. C. (2005). *Encyclopedia of statistics in behavioral sciences*. Chichester: John Wiley & Sons, Ltd.

Floyd, F. J. & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, *7*(3), 286-299.

Gable, R. K. & Wolf, M. B. (2001). *Instrument development in the affective domain: Measuring attitudes, and values in corporate and scholl settings* (Second Edition). London: Kluwer Academic Publishers.

Green, S. B., Salkind, N. J. & Akey, T. M. (1997). *Using SPSS for windows: Analyzing and understanding data*. New Jersey: Prentice Hall, Inc.

Gregorich, S. (2006). Do self report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11-3), 78-94.

Gorsuch, R. L. (1974). *Factor analysis*. (First Edition). Philadelphia: W. B. Saunders Company.

Harrington, D. (2009). *Confirmatory factor analysis* (First Edition). New York: Oxford University Press, Inc.

Jonson, J. L. & Plake, B. S. (1998). A historical comparison of validity standards and validity practices. *Educational and Psychological Measurement*. *58*(5), 736-754.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409-426.

Jöreskog, K. G. & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Lincolnwood: Scientific Software International, Inc.

Jöreskog, K. G. & Sörbom, D. (2001). *LISREL 8: User's reference guide*. Lincolnwood: Scientific Software International, Inc.

Jöreskog, K. G., Sörbom, D., Toit, M. & Toit, S. (2000). *LISREL 8: New statistical features*. Lincolnwood: Scientific Software International, Inc.

Kline, P. (2000). *The handbook of psychological testing* (Second Edition). London: Taylor & Francis Group.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (Second Edition). New York: Guilford Publications, Inc.

Lee, S. Y. & Leung, T. K. (1982). Covariance structure analysis in several populations. *Psychometrika*, 47, 297-308.

Marcoulides, G. A. & Schumacker, R. E. (1996). *Advanced structural equation modeling: ıssues and techniques* (First Edition). New Jersey: Lawrence Erlbaum Associates, Inc.

Maruyama, G. M. (1998). *Basics of structural equation modeling* (First Edition). Thousand Oaks, CA: SAGE Publications, Inc.

Meade, A. W., Michels, L. C., & Lautenschlager, G. J. (2007). Are internet and paper-and-pencil personality tests truly comparable? An experimental design measurement invariance study. *Organizational Research Methods, 10,* 322-345.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525-543.

Nunnaly, J. C. & Bernstein, I. H. (1994). *Psychometric theory* (Third Edition). New York: McGraw-Hill, Inc.

Raykov, T. & Marcoulides, G. A. (2008). An introduction to applied multivariate analysis (First Edition). New York: Taylor & Francis Group.

Rosenthal, R & Rosnow, R. L. (2008). *Essential of behavioral research* (Third Edition). New York: McGraw-Hill, Inc.

Sass, D. A., Schmitt, T. A. & Marsh H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*, 167-180.

Satorra, A. & Bentler, P. M. (2011). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika: A Journal of Quantitative Psychology*, *66*(4), 507-514.

Schmitt, N. & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*, 210-222.

Schumacker, R. E. & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling* (First Edition). New Jersey: Lawrence Erlbaum Associates, Inc.

Spini, D. (2003). Measurement equivalence of 10 value types from the Schwartz value survey across 21 countries. *Journal of Cross-Cultural Psychology*, *34*(1), 3-23.

Stapleton, C. D. (1997). Basic contepts and procedures of confirmatory factor analysis. Austin: The Annual Meeting of the Southwest Educational Research Association.

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences, 42*, 893-898.

Stevens, J. (1996). Applied multivariate statistics for social sciences (Third Edition). New Jersey: Lawrence Erlbaum Associates, Inc.

Tatlıdil, H. (1992). *Uygulamalı çok değişkenli istatistiksel analiz* (Birinci Baskı) Ankara: Engin Yayınları.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications* (First Edition). Washington: American Psychological Association.

Toit, M. & Toit, S. (2001). *Interactive LISREL: User's guide*. Lincolnwood: Scientific Software International, Inc.

Tucker, L. R. & MacCallum, R. C. (1997). Exploratory factor analysis. (Online Edition) Web: http://quantrm2.psy.ohio-state.edu/maccallum/factornew.htm adresinden 19 Şubat 2007'de alınmıştır.

Ullman, J. B. (2001). Structural equation modeling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using multivariate statistics* (pp. 653-771). Needham Heights, MA: Allyn & Bacon.

Urbina, S. (2004). *Essentials of psychological testing* (First Edition). New Jersey: Wiley & Sons, Inc.

Vandenberg, R. J. & Lance, C. E. (2000). A Review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4-70.

Westen, D. & Rosenthal, R. (2005). Improving construct validity: Cronbach, Meehl, and neurath's ship. *Psychological Assessment*, *17*(4), 409-41.

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281-324). Washington, DC: American Psychological Association.

Wu. A. D., Li, Z. & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multigroup confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, *12*(3), 1-26.

Appendix 1: LISREL Syntax Example for Test of Covariance Matrices Equality

```
SCALE I EQUALITY TEST OF COVARIANCE MATRIX (ACADEMIC MAJOR)
Group SCIENCE:
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
Latent Variables: f1-f30
Relationships:
V1 = 1*f1
V2 = 1*f2
V3 = 1*f3
V4 = 1*f4
V5 = 1*f5
V6 = 1*f6
V7 = 1*f7
V8 = 1*f8
V9 = 1*f9
V10 = 1*f10
V11 = 1*f11
V12 = 1*f12
V13 = 1*f13
V14 = 1*f14
V15 = 1*f15
V16 = 1*f16
V17 = 1*f17
V18 = 1*f18
V19 = 1*f19
V20 = 1*f20
V21 = 1*f21
V22 = 1*f22
V23 = 1*f23
V24 = 1*f24
V25 = 1*f25
V26 = 1*f26
V27 = 1*f27
V28 = 1*f28
V29 = 1*f29
V30 = 1*f30
Set the Error Variances of V1-V30 to zero
Group HEALTHCARE:
Observed Variables: V1-V30
Covariance Matrix from File HEALTHCARE.COV
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
Latent Variables: f1-f30
Group HUMANITIES:
Observed Variables: V1-V30
Covariance Matrix from File HUMANITIES.COV
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
Latent Variables: f1-f30
End of Problem
```

Appendix 2: LISREL Syntaxes Example for Test of Measurement Invariance

```
Model A
Group 1: Testing Equality Of Factor Structures
Model A: Factor Loadings, Factor Correlation, Error Variances Invariant
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Asymptotic Covariance Matrix from File SCIENCE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5
```

```
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
End of Problem
Model B
Group 1: Testing Equality Of Factor Structures
Model B: Factor Correlation and Error Variances Invariant
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Asymptotic Covariance Matrix from File SCIENCE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
End of Problem
Model C
Group 1: Testing Equality Of Factor Structures
Model C: Factor Correlation Invariant
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Asymptotic Covariance Matrix from File SCIENCE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
```

```
Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Set the Error Variances of V1-V30 free
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Set the Error Variances of V1-V30 free
End of Problem
Model D
Group 1: Testing Equality Of Factor Structures
Model D: Factor Loadings and Factor Correlation Invariant
Observed Variables: V1-V30
Covariance Matrix from File SCIENCE.COV
Asymptotic Covariance Matrix from File SCIENCE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 215
Latent Variables: Factor1 Factor2 Factor3 Factor4 Factor5
Relationships:
V2 V8 V15 V19 V23 V27=Factor1
V7 V14 V18 V24 V25 V30=Factor2
V1 V3 V4 V9 V11 V16 V22=Factor3
V5 V12 V17 V20 V29=Factor4
V10 V13 V21 V26 V28=Factor5
Group 2: Testing Equality Of Factor Correlations
Covariance Matrix from File HEALTHCARE.COV
Asymptotic Covariance Matrix from File HEALTHCARE.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 212
Set the Error Variances of V1-V30 free
Group 3: Testing Equality Of Factor Correlations
Covariance Matrix from File HUMANITIES.COV
Asymptotic Covariance Matrix from File HUMANITIES.ACM
Method of Estimation: Maximum Likelihood
Iterations: 20
Sample Size: 239
Set the Error Variances of V1-V30 free
End of Problem
```

TURKISH PRE-SERVICE EFL TEACHERS' PERCEPTIONS OF CREATIVITY

Gökçe KURT [iD]

Marmara University

gokce.kurt@marmara.edu.tr

Correspondence (only of the corresponding author)

Okan ÖNALAN [iD]

National Defense University

okanonalan@gmail.com

Biodatas

Dr. Gökçe KURT holds PhD in English Language Teaching (ELT). Her research interests include pre-service and in-service teacher education and technology integration in English language teaching and learning.

Dr. Okan ÖNALAN holds PhD in English Language Teaching (ELT) field. He currently teaches English and offers teacher-training courses at various levels.

TURKISH PRE-SERVICE EFL TEACHERS' PERCEPTIONS OF CREATIVITY

Gökçe KURT

gokcekurt@hotmail.com

Okan ÖNALAN

okanonalan@gmail.com

Abstract

Creativity is considered one of the key 21st century skills and has received an increasing amount of attention in the field of education. How teachers perceive creativity has a significant impact on their pedagogical practices to stimulate student creativity in the classroom. The present exploratory study, based on a qualitative approach, aimed to investigate pre-service English teachers' perceptions of creativity. Analysis of data coming from a semi-structured focus-group interview with eight pre-service teachers revealed their perception of creativity as being different and unique, having extraordinary ideas and being able to produce original outcomes. The participants posited that creativity was innate to some extent and environmental factors such as family environment, cultural background and school education contributed to its development later in life. They suggested using open-ended, communicative, and collaborative tasks in English classrooms to foster student creativity. As prospective teachers, they perceived themselves creative to some extent, but felt unprepared to teach creativity. Finally, they all agreed on the facilitative role of teacher education in helping PTs teach creatively and develop an understanding of teaching creativity.

Keywords: creativity, teaching creativity, pre-service teacher education, ELT.

1. Introduction

For over a century, the research on creativity has produced findings that have had significant impact on personal, social and educational domains. The concept of creativity has been investigated by researchers focusing on its psychological aspects. This line of research has attempted to understand, describe and assess the development of human creativity; and as a result, the findings have revealed that creativity is an important factor in most essential skills such as language acquisition and critical thinking (Runco, 2007). From the educational perspective, too, the creativity research has significantly reshaped educational goals, administrative matters and even the school setting as a part of 21st Century Skills.

The relationship between creativity and education works both ways. Not only does education have a vital role in stimulating learners' creative and innovative thinking (Cropley, 2001; Henessey & Amabile, 1987; Runco, 1993; Starko, 2005; Sternberg & Williams, 1996; Torrance; 1983), but creativity also contributes to learning because as learners build knowledge they utilize creative effort (Beghetto & Kaufman, 2007). In this framework, teachers are often expected to cultivate students' creative potential by helping them gain knowledge and skills related with creativity. What is more, as role models, teachers have a significant part in achieving creativity in educational settings (Berki, 2005).

From this standpoint, it is acknowledged that teachers' conceptions might foster or obstruct students' behavior since how teachers systematize their instructional practices is principally influenced by their knowledge and beliefs (Beghetto, 2006). Therefore, teachers' perceptions of creativity should be taken into account in any educational setting with an objective to develop creativity in students. In other words, it is initially important to find out how teachers conceptualize creativity in their own particular educational settings if we want to make sure that teachers want to facilitate students' creative potential and that they know how to do so (Runco, 2003).

However, in spite of the common consensus on the teachers' significant role in facilitating students' creative thinking and expression, the research that examines teacher perceptions on creativity is somehow limited. As Fryer and Collings (1991) put it, most studies have attempted to find out teachers' perceptions indirectly by measuring perceptions prior and subsequent to creativity workshops or have focused on teachers' attitudes towards the personality traits of creative individuals. What is more, there is inconsistency between the findings of such research (Westby & Dawson, 1995). Therefore, it is necessary to conduct further research focusing on in-service and prospective teachers' creativity conceptions (Nickerson, 1999) and the current study has been conducted upon this need.

2. Literature Review

In the educational framework, the word creativity is a common term, but the definitions provided by the educational and psychological researchers as well as educators are somewhat vague (Sawyer 2006a; 2006b). The term creativity may signify the process, person, product, or context; but the definitions may include one or several of these factors together, even in a contradictory manner (Taylor, 1988). Despite this elusiveness, research indicates some common themes in the definitions. Researchers and educators fundamentally concur that creativity refers to "the ability to produce work that is novel (original, unexpected) and at the same time appropriate so that the outcome meets the constraints and requirements of a task at hand" (Kaufman & Baer, 2004, p.6).

Creativity researchers also agree that creativity is commonly "a process that leads to an outcome that is novel, original, unconventional" (Ryhammar & Brolin, 1999, p.262). Sawyer (2006b) gives a broader definition by stating that creativity is the appearance of something original and appropriate. Plucker, Beghetto, and Dow (2004) assert that "creativity is the interaction among aptitude, process, and environment by which an individual or group produces a perceptible product that is both novel and useful as defined within a social context" (p. 90). Correspondingly, according to Valtanen et al., (2008), creativity requires critical and reflective thinking. Within the same framework, Dineen, Samuel, and Livesey (2005) maintain that creativity necessitates both divergent and productive thinking in order to ensure novelty as well as appropriateness.

According to the research on teachers' beliefs about creativity, it is not only researchers who have different opinions on creativity. Results of studies on teachers' views of creativity have also revealed that in the classroom setting, teachers often attempt to develop the creative potential of children when they see one (Chappell, 2007; Runco & Johnson, 2002), but they have different explanations for the term (Fleith, 2000; Fryer & Collings, 1991). For example, Diakidoy and Phtiaka (2001) point out to inconsistencies between teachers' conceptions of creativity. Also, teachers tend to have a limited view of creativity with a stereotyping attitude and they emphasize the lack of attention to creativity in teacher education (Davies, et al., 2004). What is more, according to the results reported by some researchers (Beghetto, 2006; Fasko, 2001; Runco, 2003; Westby & Dawson, 1995), although teachers generally seem to

value creativity, they hold negative attitudes and little tolerance for behaviors and attributes associated with it.

From this perspective, how teachers perceive creativity as a concept and the extent to which such perceptions align with definitions of creativity in literature proves to be important for teachers who wish to facilitate creativity in their learners (Runco & Johnson, 2001). Contemporary research on teachers' conceptions about creativity has revealed valuable outcomes about what teachers think about the nature of creativity, their conceptualization of creative individuals, and their beliefs about the kind of classrooms that promote creativity. A comprehensive review of this line of research by Andiliou and Murphy (2010), in which they analyzed peer-reviewed, empirical research studies of teachers' beliefs about creativity appearing in the published literature, suggests that the beliefs that teachers hold about creativity have a significant impact on the role they undertake in relation to creative thinking as a learning objective and on the instructional strategies they use to develop the personality characteristics required for creative behavior. They also conclude that there is still a need for further research on teachers' conceptions of creativity and their role in students' creative thinking development (Andiliou and Murphy, 2010).

Similarly, pre-service teachers' conceptions of creativity are significant since future teachers need the necessary awareness and training about creativity even before they graduate. Kampylis et al. (2009) draw attention to this issue and report that the facilitation of students' creativity is included in the teachers' role, but pre-service teachers do not feel confident and well-trained enough to fulfill this particular expectation. There are studies that investigate pre-service teachers' conceptions of creativity in various areas (e.g. Bolden et al., 2010 – mathematics). Also, recent studies within EFL (English as a foreign language) setting underline the necessity for a supportive classroom environment to promote creativity (Rao, 2014; Heathfield, 2015; Hlenschi-Stroie, 2015; Markova, 2015; Read, 2015; Wright, 2015; Wang & Kokotsaki, 2018). However, literature indicates a shortage of such studies regarding pre-service teachers' conceptions of creativity especially in language teaching field in different contexts. Thus, informed by the literature mentioned above, the present exploratory study, based on a qualitative approach, aims to investigate Turkish pre-service English teachers' perceptions of creativity. More specifically, pre-service teachers' (PTs hereafter) understanding of creativity in teaching and learning has been explored by addressing the following research questions:

1. How do Turkish PTs of English define creativity?

2. What do Turkish PTs of English think about teachability of creativity?

3. How do Turkish PTs of English perceive their own creativity?

4. How do Turkish PTs of English perceive the role of teacher education in developing creativity?

3. Method

3.1. Design, Participants and Setting

The present exploratory study adopted qualitative approach. It took place in an English Language Teaching (ELT) program of a state university in Turkey. The four-year ELT program offers basic skills courses such as Reading and Writing in English, Contextual English Grammar and Oral Communication Skills and method courses such as English Language Teaching Methods, Teaching Language Skills and Instructional Technologies and Materials Design. In the last year of the ELT Program, PTs receive two practicum courses: School Experience in English Language Education in the first term and Teaching Practice in

English Language Education in the second term. PTs are sent to different primary, secondary and high schools to observe teachers and students in the first term and to teach micro- and macro-lesson plans in the second term. At the time of the study, there were 135 PTs enrolled in the fourth year of the program. Eight PTs were selected using convenience sampling based on their willingness. Demographic characteristics of the participants were as follows: The participating PTs were all native speakers of Turkish. Four of the PTs were female and four were male. Their ages ranged from 19 to 21. None of them had any prior teaching experience. In the fall semester, they had completed the observation sessions and in the spring semester, they were doing their practicum at the assigned schools.

3.2. Data Collection and Analysis Procedure

Data for the present study came from a semi-structured focus group interview. Due to the exploratory approach of the present study, the focus group interview was preferred to gather in-depth information about PTs' perceptions of creativity. Focus-group interviews produce data and insights that cannot be easily reached without the interaction found in a group setting (Lindlof & Taylor, 2002). The face-to-face interview lasted around 50 minutes and was audio-recorded for transcription purposes. During the interview, the questions were asked in English, but PTs were free to choose the language, i.e., English or Turkish to share their opinions. The interview questions aimed to explore PTs' perceptions of the notions of creativity, teachability of creativity in the classroom, their own creativity as a future teacher, and the role of teacher education in enhancing creative ability. The transcribed data were analyzed by repeated reading to identify and report common patterns, themes and divergences within the data, as suggested by Miles and Huberman (1994).

4. Findings

4.1. PTs' Conceptions of Creativity

The interview began with PTs' definitions of creativity. They defined creativity as 'being different and unique,' 'having extraordinary ideas' and 'being able to produce new and original products.' PTs also discussed whether defining factors of creativity came from nature or nurture. There was an agreement among the PTs that creativity is innate to some extent, but can be enhanced in life. They stated that factors such as family environment, cultural background and school education influence creative behavior. PTs shared the following opinions:

I believe 30 % of creative behavior can be explained by the innate nature of creativity. 70 % is about the environmental factors.

Creativity is an innate trait, but like other abilities, it can be developed based on experience.

Children are born with creative potential. The family in which they grow up and the schools they attend to might enhance or obstruct their creativity.

I think culture also influences individuals' creativity. Some cultures value creative ideas more than others.

4.2. Teachability of Creativity

All PTs agreed on the teachability of creativity and discussed that the teacher, types of classroom tasks, the classroom environment and attitudes of school administration influence its development in the classroom. They also listed the characteristics of a creative student.

PTs identified teachers as influential figures in enhancing students' creative potential. They posited that as teachers are responsible for facilitating students' creativity development,

they should have an understanding of creativity and be knowledgeable about teaching creativity and creatively. Here are some of their comments:

The role of teachers is very important to students' creativity. Teachers should recognize the creativity in students and establish an environment that supports students' creative ability.

Teachers should believe that creativity can be taught in the classroom if suitable tasks are designed.

Teachers are responsible for teaching students skills such as questioning, analyzing, synthesizing and problem solving. These skills will foster creative performance.

Teachers may not consider themselves creative, but they can design lesson plans to promote students' creative behavior in the classroom. A wide range of teaching resources are available on the web. Teachers can adapt creative ideas to their own teaching context.

While discussing the teachability of creativity, all PTs mentioned that types of classroom tasks and activities contribute to students' creativity development too. For a foreign language classroom, PTs suggested implementing tasks that are open-ended, communicative and collaborative for the promotion of creativity. Here are some of their comments:

I believe the open-ended nature of tasks stimulates students' creativity. Students produce new, original, unique ideas. Tasks with only one possible answer might impede creativity.

Tasks based on production support creativity in the classroom. When students are asked to design a poster, create a podcast or plan a trip, they might use their creative potential.

I am doing my practicum in a private primary school. In one of the English classes I observe, a teacher asked students to design a new sport game and students came up with the idea of Mars football. I found the idea very creative. I think tasks that are different and that ask students create something new fosters students' creative potential.

When students work in groups to complete a task, they talk, exchange ideas, and help each other. They learn from each other. They also inspire each other. Together, they approach tasks and solve problems in more innovative ways.

The classroom environment was also mentioned by PTs as an important variable to enhance students' creativity. PTs believed teacher-centered classrooms where students have little or no control might impede their creativity. As one PT stated and others agreed, "students' creative thinking ability cannot be developed in a classroom environment where students are passive recipients of input." One PT also referred to the size of the classroom and said that "crowded classes are difficult to manage for creative teaching. It is easier to design creative instruction for a small number of students." As suggested by another PT and agreed by others, availability of a wide range of materials and tools also stimulates creativity.

A classroom equipped with computers with Internet connection, a projector, or a smartboard creates better opportunities for creative teaching. Students can also produce more creative products such as a podcast or a digital story in such a classroom environment.

PTs' responses to the question about the teachability of creativity also focused on the attitudes of school administration. They agreed on the importance of school support for teachers' creative pedagogical practices. The following comment of a PTs summarizes all PTs' opinions on the topic:

School administration should not force teachers to follow the assigned coursebook closely, as instructed in a teacher's manual. Rather, teachers should be encouraged to design their own methodology to foster creativity in the classroom.

Another theme that emerged during the discussion was features of a creative student. All PTs agreed that it is important for teachers to recognize these characteristics of creative students and establish a learning environment where they feel comfortable and respected. Here are their definitions of creative students:

Creative students ask many questions.

Creative students come up with unexpected, unusual, original answers or solutions.

Creative students easily get bored when they are exposed to similar materials and instruction again and again.

Creative students are usually risk takers.

Creative students are good at adapting or improving ideas and products of other people.

4.3. PTs' Perceptions of their Own Creativity

During the interview, PTs were asked whether they perceived themselves creative as teacher candidates. PTs' responses revealed that all PTs except one had considered themselves creative only to a limited extent. Only one PT said he had always considered himself creative. However, they all felt unprepared to teach creativity. Here are some of their comments:

I don't think I am really creative. This worries me because I want to be a creative teacher.

I am a creative person. I can always find creative solutions to problems in life, but I don't feel competent enough to use my creativity for teaching.

Creativity is an important skill for 21st century learning. I know I should teach my students creativity, but don't really know how.

When I was learning English, my teachers did not have a very creative way of teaching. Their practices were kind of traditional- not fostering our creativity in the classroom. This might explain why I do not find myself very creative.

4.4. PTs' Perceptions of the Role of Teacher Education

PTs had a consensus on the supportive role of teacher education in developing their creativity. They believed their creativity had increased since their entry to the program. They mentioned the courses they had received and the attitudes of the instructors as contributing to their understanding and practice of creativity. The following comments illustrate their points:

I never considered myself very creative, but I think my creativity has developed in this program. Receiving practice-based courses such as Task-based Instruction and Teaching English to Young Learners and doing my practicum at a primary school contributed to my creativity.

Being creative had never been my concern before I started this program. In the methodology courses, we are supposed to design classroom tasks, teaching materials or lesson plans. I force myself to generate original, creative ideas.

Observing a real classroom for my practicum allows me to see practices of creative teaching. Some teachers design very creative teaching tasks. It is great to see how, in a real classroom setting, students enjoy these tasks and come up with creative outcomes. I feel much inspired.

I had never thought about creativity before I came to this department. But here our instructors emphasize it as a 21ˢᵗ century skill and expect us to be creative as future teachers.

In one of the methodology courses, we were supposed to work in groups to create tasks for a specific group of EFL learners. The instructor encouraged us to generate original ideas. Nothing different came to our mind but when we talked to the instructor, she gave us inspiring examples.

The interview ended with PTs' suggestions for fostering creativity in teacher education. PTs stated that a course specifically focusing on creativity would enhance their understanding of creativity, teaching for creativity and teaching creatively. They also posited that PTs should be provided with more opportunities to observe teachers in real EFL classrooms. As they all agreed on, being introduced to a real classroom setting, observing classroom instruction of different teachers and designing lesson plans at micro- and macro-levels made PTs aware of the gap in their creative ability.

To summarize, the present study aimed to explore Turkish EFL PTs' understanding of creativity, opinions about teachability of creativity in the classroom, perceptions of their own creative ability as a future teacher, and their views on the role of teacher education in enhancing creative thinking and behavior. Eight PTs participated in the study and data came from a focused-group interview. Findings revealed that PTs perceived creativity as being different and unique, having extraordinary ideas and being able to produce original products. They believed creativity was innate to some extent and environmental factors contributed to its development later in life. All PTs agreed that creativity could be taught in the classroom as long as teachers had an understanding of creativity and implemented appropriate instruction for creativity development in a supportive classroom environment and the school had a positive attitude to its promotion in the classroom. PTs believed that open-ended, communicative tasks completed collaboratively would foster students' creative thinking in the classroom. PTs considered themselves creative to some extent, but felt unprepared to teach creativity. Finally, they all agreed on the facilitative role of teacher education in helping PTs teach creatively and develop an understanding of teaching creativity.

5. Discussion and Conclusion

There has been a considerable debate concerning the conceptualization of creativity. A widely agreed upon definition of creativity states that creativity requires both novelty and usefulness (Runco & Jaeger, 2012). Novelty or originality is crucial for creativity but is not sufficient. Creative ideas or products should also be useful or effective. In the present study, PTs' definitions revealed their limited conceptualization of creativity. PTs recognized original products as part of creativity, but could not further define that this product must be useful. PTs' definitions also focused on the product rather than the process. Cohen (1989) posits that emphasizing the product means neglecting the process of creativity, which might result in teachers' ignoring students' everyday insights for developing creative products.

PTs' discussion of creativity also focused on the question of whether creativity is an innate personality trait or something that can be developed over time. PTs agreed on the idea that all individuals have their own creativity potential supporting the views of many researchers (e.g., Feldman, Csikszentmihalyi, & Gardner, 1994; Gardner, 2011), and this potential can be developed or suppressed later in life depending on some factors such as family environment, cultural background or school education. Those factors have been shown to be critical to the development of creativity in many other studies (Fasko, 2001; Deng, Wang, & Zhao, 2016; Fleith, 2000; Niu & Sternberg, 2003).

PTs believed in teachability of creativity in the classroom and stated that creativity can be fostered through variety of tasks in a positive school environment where teachers are aware of their significant role in stimulating student creativity and supported by school administration. The role of the teacher and the school environment in developing students' creativity has been mentioned by many researchers (Feldhusen, 2005; Fleith, 2000; Kruif, McWilliam, Ridely & Wakely, 2000; Mourgues et al., 2014). Aljughaiman and Mowrer-Reynolds (2005) posited that "to foster student creativity teachers need to identify characteristics of the creative personality, recognize creative production, understand the cognitive processes used by creative students and ultimately establish an environment that promotes the students' interest" (p. 17). PTs in the present study could identify some of the characteristics of a creative student. They defined a creative student as being curious, original, social, risk-taker and energetic. The descriptors mentioned by the PTs were also used by teachers in various studies (Kampylis et al., 2009; Liu & Lin, 2014; Runco & Johnson, 2002; Sak, 2004).

In relation to teachability of creativity, PTs posited that being presented with a task or a problem in the classroom would encourage students to produce an original outcome or a solution. Their suggestion supports the view that creativity is manifested when an individual encounters a novel, ill-structured situation and attempts to define the problem at hand (Mumford, et al., 1991). For the development of creativity, PTs also emphasized creating a student-centered learning environment based on collaboration. This finding is in line with the studies discussing sociocultural and constructivist learning theories in relation to creativity development (Chan & Chan, 1999; Kampyliset et al., 2009; Myhill & Wilson, 2013; Rubenstein, McCoach, & Siegle, 2013; Sak, 2004; Zbainos & Anastasopoulou, 2012).

As future teachers, PTs perceived themselves creative to some extent but did not feel prepared to foster creativity. PTs agreed that teacher education contributed to their understanding of creativity, creative teaching, and teaching creativity, but not sufficiently. Similar concerns were reported by pre-service and in-service teachers in previous studies (Kampyliset et al., 2009; Eckhoff, 2011).

To conclude, this study aimed to contribute to the relevant literature by focusing on the perceptions of Turkish PTs' of English on creativity. As one of the key 21st century skills, creativity has been receiving an increasing attention in various fields of education including foreign/second language teaching. It is widely accepted that teachers' understanding of creativity is likely to influence their pedagogical practices to foster students' creative thinking and behavior. Teachers' beliefs on creativity are translated into their practice as long as creativity is considered an important learning goal and conditions necessary for its development are ensured to empower teachers to achieve this goal (Andiliou & Murphy, 2010). In this vein, it is important for PTs to develop adequate conceptions of creativity. Teacher education plays a significant role in providing PTs with training on the nature of creativity, the characteristics of creative students and the requirements of a classroom environment to promote creativity. Thus, pre-service English teacher education programs should be more explicit in their approach to creative teaching and teaching creativity if newly qualified teachers are to develop student creativity in the classroom.

References

Aljughaiman, A., & Mowrer-Reynolds. E. (2005). Teachers' conceptions of creativity and creative students. Journal of Creative Behavior, 39, 17- 34.

Andiliou, A., & Murphy, K. P. (2010). Examining variations among researchers' and teachers' conceptualizations of creativity: A review and synthesis of contemporary research. *Educational Research Review, 5,* 201-219.

Beghetto, R. A. (2006). Does creativity have a place in classroom discussions? Prospective teachers' response preferences. *Thinking Skills and Creativity*, *2*(1), 1–9.

Beghetto, R. A., & Kaufman, J. C. (2007). Toward a broader conception of creativity: A case for mini-c creativity. *Psychology of the Aesthetics, Creativity and the Arts, 1(2),* 73–79.

Berki, E. (2005). Mentoring as a learning process—relationships and communication. *Investigations in University Teaching and Learning*, *3*(1), 42–49.

Bolden, D.S. and Harries, A.V. and Newton, D.P. (2010) 'Pre-service primary teachers' conceptions of creativity in mathematics.', *Educational studies in mathematics. 73*(2). pp. 143-157.

Chan, D.W., & Chan, L. (1999). Implicit theories of creativity: Teachers' perceptions of student characteristics in Hong Kong. *Creativity Research Journal, 12,* 185-195.

Chappell, K. (2007). The dilemmas of teaching for creativity: Insights from expert specialist dance teachers. *Thinking Skills and Creativity, 2,* 39–56.

Cohen, L. M. (1989). A continuum of adaptive creative behaviors. *Creativity Research Journal, 2*, 169–183.

Cropley, A. J. (2001). *Creativity in education & learning: A guide for teachers and educators*. London: Kogan Page.

Davies, D., Howe, A., Fasciato, M., & Rogers, M. (2004). How do trainee teachers understand creativity? In Proceedings of the DATA international research conference: Creativity and innovation 2004 Wallesbourne, UK.

Deng, L., Wang, L. & Zhao, Y. (2016) How creativity was affected by environmental factors and individual characteristics: A cross-cultural comparison perspective. *Creativity Research Journal, 28(3*), 357-366.

Diakidoy, I.-A., & Phtiaka, H. (2001). Teachers' beliefs about creativity. In S. S. Nagel (Ed.), Handbook of policy creativity: Creativity from diverse perspectives (pp. 13–32). Huntington, NY: Nova Science.

Eckhoff, A. (2011). Creativity in the early childhood classroom: perspectives of preservice teachers. *Journal of Early Childhood Teacher Education, 32(3*), 240-255.

Fasko, D. (2001). Education and creativity. *Creativity Research Journal, 13*(3/4), 317–327.

Feldhusen, J. F. (2005). Giftedness, Talent, Expertise, and Creative Achievement. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 64-79). New York, NY, US: Cambridge University Press.

Feldman, D. H., Csikszentmihalyi, M., & Gardner, H. (1994). *Changing the world: A framework for the study of creativity.* Westport, CT, US: Praeger Publishers/Greenwood Publishing Group.

Fleith, D. S. (2000). Teacher and student perceptions of creativity in the classroom environment. *Roeper Review, 22(3),* 148–153.

Fryer, M., & Collings, J. A. (1991). British teachers' views of creativity. *The Journal of Creative Behavior, 25* (1), 75-81.

Gardner, H. (2011). Intelligence, Creativity, Ethics: Reflections on My Evolving Research Interests. *Gifted Child Quarterly, 55*(4), pp. 302 – 304.

Heathfield, D. (2015). Personal and creative storytelling: Telling our stories. In Maley, A., & Peachey, N. (Eds.). *Creativity in the English language classroom (pp.44-50).* London: British Council.

Henessey, B.A., & Amabile, T.M. (1987). *Creativity and learning.* Washington DC: National Education Association.

Hlenschi-Stroie, V. (2015). Drama and creative writing: A blended tool. In Maley, A., & Peachey, N. (Eds.). *Creativity in the English language classroom (pp.158-164).* London: British Council.

Kaufman, J. C., & Baer, J. (2004). Hawking's Haiku, Madonna's Math: Why it is hard to be creative in every room of the house. In R. J. Sternberg, E. L. Grigorenko, & J. L. Singer (Eds.), Creativity: From potential to realization (3–20). Washington, DC: American Psychological Association.

Kampylis, Panagiotis & Berki, Eleni & Saariluoma, Pertti. (2009). In-service and prospective teachers' conceptions of creativity. *Thinking Skills and Creativity. 4.* 15-29.

Kruif, R. E. L., McWilliam, R. A., Ridley, S. M., & Wakely, M. B. (2000). Classification of teachers' interaction behaviors in early childhood classrooms. *Early Childhood Research Quarterly, 15*(2), 247-268.

Lindlof, T. R. & Taylor, B. C. (2002). *Qualitative communication research methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Liu, S. C., & Lin, H.S. (2014). Primary teachers' beliefs about scientific creativity in the classroom context. *International Journal of Science Education, 36*(10), 1551-1567

Markova, Z. (2015). A journey towards creativity: A case study of three primary classes in a Bulgarian state school. In Maley, A., & Peachey, N. (Eds.). *Creativity in the English language classroom (pp.165-172).* London: British Council.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook.* Thousand Oaks: Sage Publications.

Mourgues, C. V., Tan, M., Hein, S., Al-Harbi, K., Aljughaiman, A., and Grigorenko, E. L. (2016). The relationship between analytical and creative cognitive skills from middle childhood to adolescence: testing the threshold theory in the kingdom of Saudi Arabia. *Learner Individual Differences, 52,* 137–147.

Mumford, M. D., Mobley, M. I., Uhlman, C. E., Reiter-Palmon, R., & Doares, L. M. (1991). Process analytic models of creative capacities. *Creativity Research Journal, 4,* 91-122.

MyHill, D. & Wilson, A. (2013). Playing it safe: Teachers' views of creativity in poetry writing. *Thinking Skills and Creativity, 10*, pp. 101-111.

Nickerson, R. (1999). Enhancing creativity. In R. J. Sternberg (Ed.), *Handbook of creativity* (pp. 189–212). New York: Cambridge University Press.

Niu, W., & Sternberg, R. J. (2003). Societal and school influences on student creativity: The case of China. *Psychology in the Schools, 40*(1), 103-114.

Plucker, J. A., Beghetto, R. A., & Dow, G. T. (2004). Why isn't creativity more important to educational psychologists? Potential, pitfalls, and future directions in creativity research. *Educational Psychologist, 39(2),* 83–96.

Rao, L. (2014). Creative education in primary school English (Xiaoxue yingyu chuangxin jiaoyu). *New Curriculum, 6, 76.*

Read, C. (2015). Seven pillars of creativity in primary ELT. In Maley, A., & Peachey, N. (Eds.). *Creativity in the English language classroom (pp.29-36).* London: British Council.

Rubenstein, L.D., McCoach, D.B., & Siegle, D. (2013). Teaching for Creativity Scales: An instrument to examine teachers' perceptions of factors that allow for the teaching of creativity. *Creativity Research Journal, 25*, 324–334

Runco, M. A. (1993). *Creativity as an educational objective for disadvantaged children.* Storrs, CT: The National Research Center on the Gifted and Talented.

Runco, M. A. (2003). Education for creative potential. *Scandinavian Journal of Educational Research*, *47*(3), 317–324.

Runco, M. A. (2007). *Creativity, theories and themes: Research, development and practice.* London, UK: Elsevier Academic Press.

Runco, M. A., & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, 24, 92-96.

Runco, M. A., Johnson, D. J., & Bear, P. K. (1993). Parents' and teachers' implicit theories of children's creativity. *Child Study Journal, 23(2),* 91–113.

Runco, M. A., & Johnson, D. J. (2002). Parents' and teachers' implicit theories of children's creativity: A cross-cultural perspective. *Creativity Research Journal*, *14*(3/4), 427–438.

Ryhammar, L., & Brolin, C. (1999). Creativity research: Historical considerations and main lines of development. *Scandinavian Journal of Educational Research, 43(3),* 259–273.

Sak, U. (2004). About Creativity, Giftedness and Teaching the Creativity Gifted in the Classroom. *Roeper Review, 26*(4), 216-222.

Sawyer, R. K. (2006a). Educating for innovation. *Thinking Skills and Creativity*, *1*(1), 41–48.

Sawyer, R. K. (2006b). *Explaining creativity: The science of human innovation*. Oxford\New York: Oxford University Press.

Starko, A. J. (2005). *Creativity in the classroom: Schools of curious delight* (3rd ed.). Mahwah, NJ: L. Erlbaum Associates.

Sternberg, R.J., & Williams, W.M. (1996). *How to develop student creativity.* Alexandria, VA: Association for Supervision and Curriculum Development.

Taylor, C. W. (1988). Various approaches to and definitions of creativity. In R. J. Sternberg (Ed.), *The nature of creativity—Contemporary psychological perspectives*. Cambridge: Cambridge University Press.

Torrance, E.P. (1983). *Creativity in the classroom.* Washington, DC: National Education Association.

Valtanen, J., Berki, E., Kampylis, P., & Theodorakopoulou, M. (2008). Manifold thinking and distributed problem-based learning: Is there potential for ICT support? In Proceedings of the IADIS international conference e-learning 2008 Amsterdam, Netherlands.

Wang L., & Kokotsaki D. (2018). Primary school teachers' conceptions of creativity in teaching English as a foreign language (EFL) in China, *Thinking Skills and Creativity.* https://doi.org/10.1016/j.tsc.2018.06.002

Westby, E. L., & Dawson, V. L. (1995). Creativity: Asset of burden in the classroom? *Creativity Research Journal*, *8*(1), 1–11.

Wright, A. (2015). Medium: Companion or slave? In Maley, A., & Peachey, N. (Eds.). *Creativity in the English language classroom (pp.14-23).* London: British Council.

Zbainos, D., Anastasopoulou, A. (2012). Creativity in Greek music curricula and pedagogy: An investigation of Greek music teachers' perceptions. *Creative Education, 3*(1), 55–60

TOWARDS INTERCULTURAL COMPETENCE: MODELS AND FRAMEWORKS FOR DEVELOPING ESOL LEARNERS' INTERCULTURAL COMPETENCE IN IRELAND

György Nagy

University of Limerick, Ireland

gyorgy.nagy@ul.ie

György Nagy is a Hungarian PhD student in Teaching English to Speakers of Other Languages (TESOL). His research explores the potential in materials for developing learners' intercultural competence in an Irish context. György has been active in TESOL for many years.

TOWARDS INTERCULTURAL COMPETENCE: MODELS AND FRAMEWORKS FOR DEVELOPING ESOL LEARNERS' INTERCULTURAL COMPETENCE IN IRELAND

György Nagy

gyorgy.nagy@ul.ie

Abstract

Without the study of culture, second language acquisition is not complete (Kramsch, 1993). Teaching about culture raises the learners' awareness of the target culture and compares it with the home culture, providing an intercultural competence (Ibid). Due to the growing number of immigrants in Ireland (CSO Ireland, 2017), there is also an increasing number of immigrant learners in the English for Speakers of Other Languages (ESOL) providers in Ireland. The successful integration of these learners into Irish society depends on the successful development of their intercultural competence. This paper aims to provide insight into the models and frameworks for developing intercultural competence through the use of English language teaching materials in an Irish context. The paper involves presenting four models-based frameworks. The frameworks consider increasing content knowledge, sharpening mental skills, fostering attitudinal skills and increasing awareness which constitute the four fundamental components of intercultural competence. The paper endeavours to contribute to the vibrant global conversation among professionals about how to develop intercultural competence. Most significantly, however, it attempts to support teachers in incorporating cultural elements into their teaching materials effectively and appropriately.

Keywords: English for Speakers of Other Languages, Irish culture, intercultural competence, teaching materials

1. Introduction

Modern migratory tendencies in Europe and worldwide have led to annual increases in immigration into Ireland (CSO Ireland, 2017) with a projection that in the future, Brexit will intensify the impact on Ireland's demography (Varadkar, 2017). This trend will inevitably result in more and more immigrant learners in the English for Speakers of Other Languages (ESOL) classrooms throughout the country. The successful integration of these learners into Irish society depends on the successful development of their intercultural competence and this, in turn, relies on the teaching about Irish culture. Based on a definition of culture from G. Hofstede, G. J. Hofstede and M. Minkov (2010), Irish culture is the way members of a group of Irish people, living in Ireland, do things and the way these members see the things they do. Drawing on this conceptualisation of culture, intercultural competence for immigrant ESOL learners in Ireland is the ability to do, see and even feel things the way the Irish do. This competence is essential to function well in a society that is new to a person. Functioning well refers to functioning *effectively* and functioning *appropriately* (Fantini & Tirmizi, 2006, p. 12, emphasis in original). Functioning effectively means achieving results within Irish society and function appropriately means achieving the results in a way which is suitable and right in the eyes of the Irish.

This study looks at how to develop immigrant learners' intercultural competence in an Irish ESOL classroom. Intercultural competence pertains to development through the

components within intercultural competence. Using the models from Byram (1997), Byram, Gribkova and Starkey (2002), Deardorff (2006) and Fantini (2009), there are four components, or dimensions, of intercultural competence. They are content knowledge, mental skills, attitudinal skills and awareness (see Figure 1). *Content knowledge* is knowledge of acquired cultural information; *mental skills* is the ability to process cultural information; *attitudinal skills* is the ability to value cultural information; *awareness* is a perception of learning in the domains of content knowledge, mental skills and attitudinal skills.
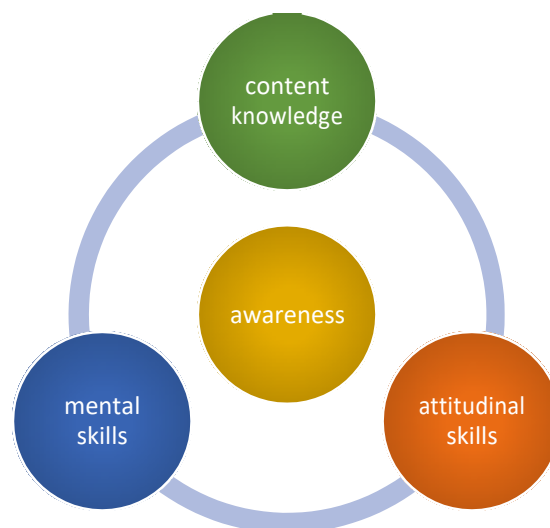
Figure 1. Components of intercultural competence

The ESOL teacher's most utilised classroom tool is materials. Based on Mishan's (2005) definition, language teaching materials are *texts* which can be either written or audio, where *visuals* can be still or moving, and *tasks* are based on the texts and visuals.

Using the concept of ESOL materials and the four components of intercultural competence in an ESOL classroom, this paper advocates the use of the four model-based frameworks. These frameworks are designed to increase the potential for deepening content knowledge through texts and visuals, sharpening mental skills and fostering attitudinal skills through tasks and increasing awareness through the material as a whole. Each framework is intended to help ESOL teachers incorporate culture into their own-devised teaching materials or to evaluate their existing teaching materials and amend them as necessary. In this way, teachers can provide a more significant potential for the development of their learners' intercultural competence in the language classroom. As the research that utilises these frameworks is being carried out in Ireland, the frameworks are applied to an Irish context.

2. Deepening content knowledge

The framework for deepening content knowledge through texts and visuals relies on two models. They are termed as 'the pentagon of culture' and 'the fields of culture'. This framework is designed to increase the potential in the texts and visuals to *acquire* cultural information in a *comprehensive* and *detailed* way.

2.1 Pentagon of culture

The pentagon of culture is the expansion of the three Ps (products, practices, perspectives) approach to culture (National Standards in Foreign Language Education Project, 2015). It consists of five broad areas: geography, people, products, practices and perspectives (see Figure 2).
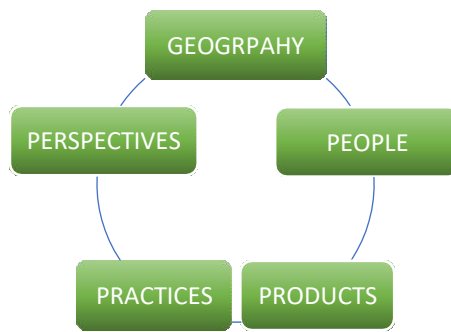
Figure 2. The pentagon of culture

By addressing these five areas, a comprehensive description of any cultural item can be ensured. As comparison lies at the heart of cultural learning, the pentagon of culture is an excellent tool for comprehensively comparing two cultures. Figure 3 shows an example of comparing 'tea' in Ireland with 'tea' in Hungary using the pentagon of culture.

Figure 3. Comparing 'tea' in Ireland with 'tea' in Hungary using the pentagon of culture

2.2 Fields of culture

The five broad areas of the pentagon of culture can further be divided into several 'fields'. The fields of culture are the expansion of the minimum areas of cultural content (Byram & Morgan, 1994) in teaching materials (see Table 1). Each field has substantial sub-fields, for instance, cultural heritage includes dance, music and literature; social identity involves age, gender, socio-economic status, language use. (See Table 2 for more details). By addressing the fields of culture, the opportunity for a detailed description of any cultural item is ensured.

Table 1. *The fields of culture*

geography	social identity	rites of passage
history	social groups	beliefs
institutions	social interactions	norms
cultural heritage	behaviours	values
physical items	socialisation	meanings

national identity	life cycles	generalisations

2.3 Framework for deepening content knowledge

The framework for deepening content knowledge through texts and visuals consists of two main parts. Firstly, the five broad areas of the pentagon of culture which ensures a comprehensive description of cultural items. Secondly, the fields of culture with their sub-fields assigned to the relevant areas of the pentagon of culture which ensures a detailed description of cultural elements (see Table 2).

Table 2. *Framework for deepening content knowledge through texts and visuals*

□ Geography	□ People	□ Products	□ Practices	□ Perspectives
natural [ ] landforms [ ] water bodies [ ] climate [ ] weather [ ] animals [ ] plants human [ ] countries [ ] counties [ ] cities, towns [ ] villages, settlements [ ] population	[ ] national identity social identity [ ] race [ ] gender [ ] age [ ] ethnicity [ ] socioeconomic status [ ] political affiliation [ ] religious affiliation [ ] language use social groups [ ] membership groups [ ] reference groups	tangible [ ] cultural heritage [ ] physical items intangible [ ] cultural heritage [ ] history [ ] language [ ] institutions [ ] life cycles [ ] other non-physical items	social interactions [ ] verbal [ ] non-verbal socialisation [ ] family [ ] school [ ] work [ ] peer group behaviours [ ] routines [ ] manners rites of passage [ ] ceremonies [ ] rituals	[ ] beliefs [ ] values [ ] norms [ ] meanings [ ] generalisations

The more complete the table is, the stronger the opportunity is to acquire cultural information in a comprehensive and detailed way as well as to increase awareness.

2.3.1 An example framework for deepening content knowledge

Text

The Claddagh Ring merges the basic 'clasped hand design' of Fede Rings with a heart and a crown signifying Love, Friendship and Loyalty or Fidelity. These rings were kept with great pride as family heirlooms, which were passed lovingly down from mother to daughter on her wedding day. The design has now become very popular internationally, its spread being helped by the vast exodus from Ireland during the Great Famine in 1845-1849. The simplicity of the ring and the symbolism it conveys make it the perfect gift for a friend or loved one. (Excerpt from Unit 18 in Jordan, M. (2005) *Learning English in Ireland,* Dublin: Celtic Publications.)

Visual

Table 3. *Examination of a text and a visual about an Irish cultural artefact*

√ Geography	√ People	√ Products	√ Practices	√ Perspectives

natural
 [ ] landforms
 [ ] water bodies
 [ ] climate
 [ ] weather
 [ ] animals
 [ ] plants
human
 [√] countries
 [ ] counties
 [ ] cities, towns
 [ ] villages,
   settlements
 [ ] population
	[√] national
	identity
	social identity
	 [ ] race
	 [√] gender
	 [√] age
	 [ ] ethnicity
	 [ ] socio-
	   economic
	   status
	 [ ] political
	   affiliation
	 [ ] religious
	   affiliation
	 [ ] language
	   use
	social groups
	 [√] membership
	   groups
	 [ ] reference
	   groups
		tangible
		 [√] cultural
		   heritage
		 [√] physical
		   items
		intangible
		[√] cultural
		   heritage
		 [√] history
		 [ ] language
		 [ ] institutions
		 [√] life cycles
		 [ ] other non-
		   physical
		   items
			social interactions
			 [ ] verbal
			 [√] non-verbal
			socialisation
			[√] family
			 [ ] school
			 [ ] work
			 [ ] peer group
			behaviours
			 [√] routines
			 [ ] manners
			rites of passage
			 [√] ceremonies
			 [ ] rituals
				[ ] beliefs
				[√] values
				[√] norms
				[√] meanings
				[√] generalisations

The example text and visual deal with all five areas of the pentagon of culture: geography, people, products, practices and perspectives. Also, many of fields and sub-fields of culture are considered within this analysis. Country, national identity, gender, age, membership groups, tangible and intangible cultural heritage, physical items, history, life cycles, non-verbal social interactions, family socialisation, routines, ceremonies, values, norms, meanings and generalisations. Consequently, the conclusion is that the example text and visual have great potential to acquire cultural information in a comprehensive and detailed way as well as to increase awareness of Irish culture (see Table 3).

3. Sharpening mental skills

The framework for sharpening mental skills through tasks comprises two models incorporated in the comparative approach to cultural learning. The two models are the previously presented pentagon of culture (see 2.1) and the cognitive learning ladder. This framework is designed to increase the potential in tasks to *process* cultural information in a *complex*, *comprehensive* and *comparative* way.

3.1 Cognitive learning ladder

The cognitive learning ladder is the taxonomy of cognitive learning objectives in educational settings (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths & Wittrock, 2001; Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) (see Figure 4). It is symbolised as a ladder to be climbed, step by step, to reach the highest step 'safely'. It relies on each step necessarily including the steps that are below it. For example, analysing cultural information (step 4) is not possible without being able to recall (step 1), comprehend (step 2)

and apply (step 3) the information to be analysed. The higher position the cognitive operation of the mind reaches on the ladder, the stronger the possibility is to process cultural information complexly.
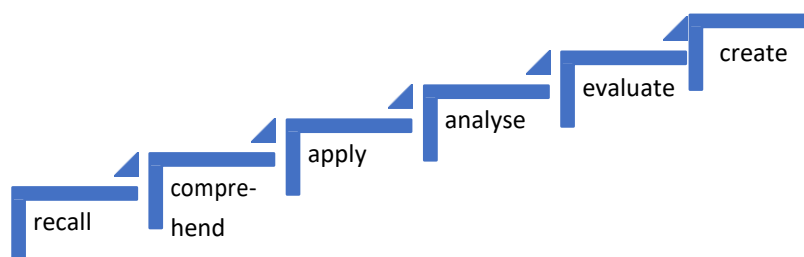

Figure 4. The cognitive learning ladder

Learning about the Claddagh Ring (see 2.3.1) may be used an example to demonstrate each step of the cognitive learning ladder. 'Recall' (step 1) is when the learner recognises the Claddagh Ring. 'Comprehend' (step 2) is when the learner understands the symbols and the uses of the ring. 'Apply' (step 3) is when the learner gives examples of the uses of the ring in real life situations. 'Analyse' (step 4) is when the learner examines the symbols, the history or the uses of the ring. 'Evaluate' (step 5) is when the learner judges and critiques the symbols and the uses of the ring. 'Create' (step 6) is when the learner designs his or her ring with a similar meaning to that of the Claddagh Ring.

3.2 Framework for sharpening mental skills

The framework for sharpening mental skills through tasks comprises three main components. The components may be summarised as firstly, the steps of the cognitive learning ladder from the lowest to the highest to ensure complexity in the tasks. Secondly, the five areas of the pentagon of culture to provide comprehensiveness in the tasks. Finally, references to Irish cultural information and information from the learner's own culture to ensure comparison in the tasks (see Table 4).

Table 4. *Framework for sharpening mental skills through tasks*

MENTAL SKILLS IN		recall comprehend	apply analyse	evaluate create
Irish	geography			
Learner's				
Irish	people			
Learner's				
Irish	products			
Learner's				
Irish	practices			
Learner's				
Irish	perspectives			
Learner's				

The more complete the table is, the stronger the opportunity is to process cultural information in a complex, comprehensive and comparative way as well as to increase awareness.

3.2.1 An example framework for sharpening mental skills

Tasks

Compare (Task 1) the Claddagh Ring with a similar traditional artefact for personal use from your own country. Think of (Task 2) similarities and differences, for example: who uses it, when, where and how, and how the people in your country view it.

Table 5. *Examination of tasks on comparing cultural artefacts*

MENTAL SKILLS IN		recall	comprehend	apply	analyse	evaluate	create
Irish	geography					Task 1	Task 2
Learner's	geography					Task 1	Task 2
Irish	people					Task 1	Task 2
Learner's	people					Task 1	Task 2
Irish	products					Task 1	Task 2
Learner's	products					Task 1	Task 2
Irish	practices					Task 1	Task 2
Learner's	practices					Task 1	Task 2
Irish	perspectives					Task 1	Task 2
Learner's	perspectives					Task 1	Task 2

Comparing, Task 1 at the level of 'evaluate', has the potential to provide the learner with an opportunity to judge and critique differences between the Claddagh Ring and a similar artefact from his or her country using all the aspects of the pentagon of culture. Also, the task naturally involves the lower-level cognitive operations from 'recall' to 'apply'. Thinking of differences and similarities, Task 2 at the level of 'create', has the potential to generate ideas and put them together in a systematic way. Task 2 also includes the lower-level cognitive operations from 'recall' to 'evaluate'. It can be concluded that the two tasks have great potential to process cultural information in a complex, comprehensive and detailed way as well as to increase awareness of the Irish and the learner's self (see Table 5).

4. Fostering attitudinal skills

The framework for fostering attitudinal skills through tasks derives two models incorporated in the comparative approach to cultural learning. The two models are the previously presented pentagon of culture (see 2.1) and the affective learning ladder. The application of this framework supports increasing the potential in tasks to *value* cultural information in a *complex*, *comprehensive* and *comparative* way.

4.1 Affective learning ladder

The affective learning ladder is the taxonomy of affective learning objectives in educational settings (Anderson et al., 2001; Krathwohl, Bloom, & Masia, 1964) (see Figure 5). It works in the same way as the previously introduced cognitive learning ladder (see 3.1); that is, each step naturally includes all the lower steps. For instance, valuing cultural information (step 4) is not possible without being able to receive (step 1) the information and respond (step 2) to it first. The higher position the affective operation reaches on the ladder, the stronger the possibility is to value cultural information in a complex way.
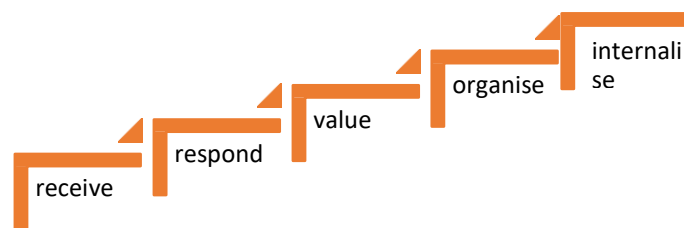
Figure 5. The affective learning ladder

Learning about the Claddagh Ring (see 2.3.1) is taken again as an example to demonstrate each step of the affective learning ladder. 'Receive' (step 1) is when the learner is willing to discover the Claddagh Ring – this is motivation. 'Respond' (step 2) is when the learner reflects positively on the symbols and the uses of the ring – this is interest. 'Value' (step 3) is when the learner accepts the meanings and uses of the ring – this is appreciation. 'Organise' (step 4) is when the learner ranks the acquired new values about the ring and integrates them into his or her old value system – this is the development of a life philosophy which might be the highest step to be reached in an ESOL classroom. 'Internalise' (step 5) is when the learner lives consistently according to the values of the Claddagh Ring – this is an adjustment.

4.2 Framework for fostering attitudinal skills

The framework for fostering attitudinal skills through tasks consists of three main components. These are the steps of the affective learning ladder from the lowest to the highest to provide complexity in the tasks. Secondly, the five areas of the pentagon of culture to ensure comprehensiveness in the tasks. Finally, references to Irish cultural information and information from the learner's own culture to ensure comparison in the tasks (see Table 6).

Table 6. *Framework for fostering attitudinal skills through tasks*

ATTITUDINAL SKILLS IN		receive	respond	value	organise	internalise
Irish	geography					
Learner's						
Irish	people					
Learner's						
Irish	products					
Learner's						
Irish	practices					
Learner's						
Irish	perspectives					
Learner's						

The more complete the table is, the stronger the opportunity is to value cultural information in a complex, comprehensive and comparative way as well as to increase awareness.

4.2.1 Example framework for fostering attitudinal skills

Tasks

Compare (Task 1) the Claddagh Ring with a similar traditional artefact for personal use from your own country. Think of (Task 2) similarities and differences, for example: who uses it, when, where and how, and how the people in your country view it.

Table 7. *Examination of tasks on comparing cultural artefacts*

ATTITUDINAL SKILLS IN		receive	respond	value	organise	internalise
Irish	geography			Task 2	Task 1	
Learner's	geography			Task 2	Task 1	
Irish	people			Task 2	Task 1	
Learner's	people			Task 2	Task 1	
Irish	products			Task 2	Task 1	
Learner's	products			Task 2	Task 1	
Irish	practices			Task 2	Task 1	
Learner's	practices			Task 2	Task 1	
Irish	perspectives			Task 2	Task 1	
Learner's	perspectives			Task 2	Task 1	

Comparing, Task 1 at the level of 'organise', has the potential to provide the learner with an opportunity to rank and internalise the moral values connected with the Claddagh Ring and a similar artefact from his or her country looking at all the aspects of the pentagon of culture. The task also involves the lower-level affective operations from 'receive' to 'value. Thinking of differences and similarities, Task 2 at the level of 'value', has the potential for the appreciation of the moral values of the two cultural artefacts. Task 2 also includes the lower-level affective operations from 'recall' to' respond'. It can be concluded that the two tasks have great potential to value cultural information in a complex, comprehensive and detailed way as well as to increase awareness of the Irish and the learner's self (see Table 7).

5. Increasing awareness

The framework for increasing awareness of the Irish and the learner's self is based on the findings noted in the frameworks for deepening content knowledge (see 2.3), sharpening mental skills (see 3.2) and fostering attitudinal skills (see 4.2) incorporating a comparative approach to cultural learning. It is designed to increase the potential in the material as a whole for perceptions of *acquiring*, *processing* and *valuing* cultural information.

5.1 Framework for increasing awareness

This very simple, tick-box framework comprises a column for 'awareness of the Irish' and a column for 'awareness of learner's self' with three rows for the findings recorded in the frameworks for content knowledge, mental skills and attitudinal skills domains of intercultural competence (see Table 8). These findings need to be looked at carefully and then, the relevant boxes of the framework need to be ticked if they are present.

Table 8. *Framework for increasing awareness through the material as a whole*

AWARENESS THROUGH	awareness of the Irish	awareness of learner's self
content knowledge		
mental skills		
attitudinal skills		

5.1.1 An example framework for increasing awareness

Table 9. *Examination of a material on cultural artefacts*

AWARENESS THROUGH	awareness of the Irish	awareness of learner's self
content knowledge	√	
mental skills	√	√
attitudinal skills	√	√

The text and the visual in the example (see 2.3.1) address an Irish cultural item (the Claddagh Ring) only but the tasks (see 3.2.1 and 4.2.1) address both Irish cultural information and information from the learner's own culture in the domains of cognitive (mental) and affective (attitudinal) learning. Thus, it can be concluded that the material, when considered as a whole, has great potential for promoting perceptions of content knowledge, mental skills and attitudinal skills (see Table 9).

6. Conclusion

Developing immigrant learners' intercultural competence in an ESOL classroom in Ireland is crucial to their successful integration into Irish society. This paper has shown how learners' intercultural competence is developed by improving the four components of intercultural competence with the help of language teaching materials using four practical frameworks. (1) The framework for deepening content knowledge helps to increase the potential in the texts and visuals to acquire cultural information in a comprehensive and detailed way. (2) The framework for sharpening mental skills helps to increase the potential in the tasks to process cultural information in a complex, comprehensive and comparative way. (3) The framework for fostering attitudinal skills helps to increase the potential in the tasks to value cultural information in a complex, comprehensive and comparative way. (4) The framework for increasing awareness of the Irish and the learner's self helps to increase the potential in the material as a whole for perceptions of the learnings in the domains of content knowledge, mental skills and attitudinal skills (see Figure 6).
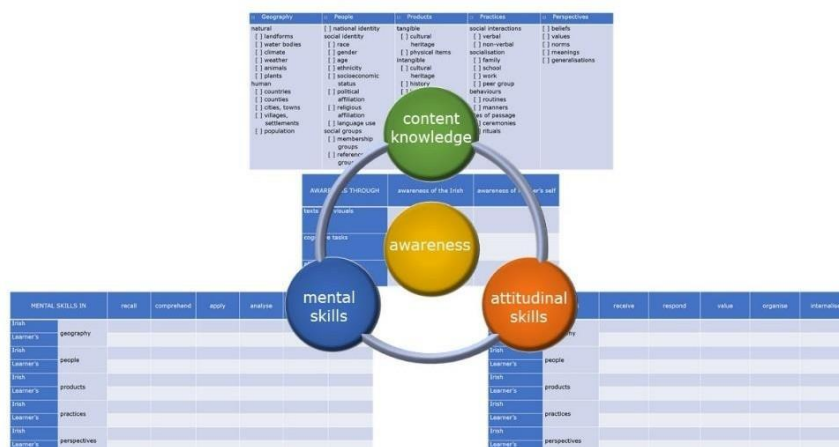

Figure 6. Frameworks for developing intercultural competence

The four frameworks are designed to help to increase the potential in language teaching materials to develop the four components of intercultural competence, and by this,

intercultural competence itself, not only in an ESOL classroom in Ireland but also in any other language classroom around the world.

References

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J. & Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives (Complete edition)*. (L. W. Anderson & D. R. Krathwohl, Eds.). New York: Longman.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* (B. S. Bloom, Ed.). New York: David McKay.

Byram, M. (1997). *Teaching and Assessing Intercultural Communicative Competence*. Clevedon: Multilingual Matters.

Byram, M., Gribkova, B., & Starkey, H. (2002). *Developing the Intercultural Dimension in Language Teaching a Practical Introduction for Teachers*. Strasbourg: Council of Europe.

Byram, M., & Morgan, C. (1994). *Teaching and Learning Language and Culture*. Clevedon: Multilingual Matters.

CSO Ireland. (2017). Population and Migration Estimates April 2017. Retrieved May 30, 2018, from https://www.cso.ie/en/releasesandpublications/er/pme/populationandmigrationestimatesapril2017/

Deardorff, D. K. (2006). The Identification and Assessment of Intercultural Competence as a Student Outcome of Internationalization at Institutions of Higher Education in the United States. *Journal of Studies in International Education*, (10), 241–266.

Fantini, A. (2009). A central concern: developing intercultural competence. *Journal of Psychology*, *37*(1), 54–61.

Fantini, A., & Tirmizi, A. (2006). Exploring and Assessing Intercultural Competence. *World Learning Publications*. Paper 1.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and Organizations: Software of the Mind*. New York: McGraw-Hill.

Jordan, M. (2005). *Learning English in Ireland*. Dublin: Celtic Publications.

Kramsch, C. (1993). *Teaching language along the cultural faultline*. In Kramsch, C. (Ed.), *Context and Culture in Language Teaching* (pp. 205-232). Oxford: Oxford University Press.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives, the Classification of educational goals. Handbook II: Affective domain*. New York: David McKay Co. Inc.

Mishan, F. (2005). *Designing authenticity into language learning materials.* Bristol: Intellect Books.

National Standards in Foreign Language Education Project. (2015). *World-Readiness Standards for Learning Languages* (4th ed.). Alexandria, VA: Author.

Varadkar, L. (2017). *National Risk Assessment 2017: Overview of Strategic Risks*. Dublin: Department of the Taoiseach.

THE ROLE OF GOAL SETTING IN METACOGNITIVE AWARENESS AS A SELF-REGULATORY BEHAVIOR IN FOREIGN LANGUAGE LEARNING

Nihan Bursalı [ID]

Ufuk University

nihanbursali@gmail.com

Hüseyin Öz [ID]

Hacettepe University

hoz@hacettepe.edu.tr

Nihan Bursalı majored English Language Teaching at Hacettepe University and also continues her studies in there. She started her carrier as an instructor and she is currently working as a research assistant at Ufuk University in English Language Teaching Department.

Hüseyin Öz is an associate professor of applied linguistics and English language teaching at Hacettepe University. He received his MA degree from Middle East Technical University and his PhD degree in Linguistics from Hacettepe University, where he teaches undergraduate and graduate courses in language teaching methods and approaches, research methods, linguistics, language testing, and technology enhanced language learning. He has published widely in various refereed journals and presented papers in national and international conferences. He has also served on the editorial boards of several national and international publications and is currently the associate and managing editor of Eurasian Journal of Applied Linguistics.

THE ROLE OF GOAL SETTING IN METACOGNITIVE AWARENESS AS A SELF-REGULATORY BEHAVIOR IN FOREIGN LANGUAGE LEARNING

Nihan Bursalı

nihanbursali@gmail.com

Hüseyin Öz

hoz@hacettepe.edu.tr

Abstract

Metacognition, an awareness of one's own cognitive processes in learning, is a crucial component of self-regulatory behavior that facilitates successful language learning. Therefore, the current study sought to find out the role of different types of goals in participants' metacognitive awareness. A total of 118 university students enrolled in an English Language Teaching program at a major state university voluntarily participated in the study. Data were gathered using Metacognitive Awareness Inventory (MAI) and Goals Inventory. Findings of descriptive statistics revealed that 48.3% of the participants had high, 28% had moderate, and 23.7% had low metacognitive awareness. Besides, significant correlation was found between mastery goals and metacognitive awareness, supporting the findings of the previous studies. These results revealed the importance of goal setting in metacognition and can be useful for practitioners to include in goal orientation to the curriculum by emphasizing the necessity of learner engagement, agency, and self-regulation for successful language learning process.

Keywords: EFL, goals, metacognition, metacognitive awareness, self-regulatory behavior

1. Introduction

Even though the terms are used interchangeably from time to time, self-regulation and metacognition do not refer to the same thing. In fact, self-regulation involves metacognition along with many other things such as "goal setting, monitoring and evaluating one's actions" (Williams, Mercer, & Ryan, 2015, p. 130). Self-regulatory learners know their own weaknesses and strengths; thus, they know themselves. Furthermore, they are good agents, they can set goals, motivate themselves, and evaluate their own learning process. Eventually, successful language learning occurs. Therefore, facilitating self-regulation has a crucial place in language learning process. The present study sought to find out the role of different types of goals, namely mastery goals and performance goals, in metacognitive awareness which is a vital component of self-regulation. In the literature, there are several related studies; however, the issue deserves more attention (Coutinho, 2007; Gul & Shehzad, 2012; Kharazi, Ezhehei, Ghazi Tabatabaei, & Kareshki, 2008; Schmidt & Ford, 2003; Zafarmand, Ghanizadeh, & Akbari, 2014).

2. Literature Review

1.1. Metacognition

Metacognition is an important element of self-regulatory behavior and could be associated with one's abilities to plan and control his or her own learning process. In other words, it can be considered as an awareness on learning processes, especially cognitive ones. Metacognitive strategies help learners have an outsider perspective on their learning process and analyze if the strategies they use are sufficient or not. On the other hand, cognitive strategies succinctly are mental processes that help learners process information that they need to learn. Therefore, while cognition refers to thinking, metacognition could be associated with thinking about thinking (Öz, 2014, 2015; Williams et al., 2015). For example, when a language learner thinks strategically to complete a task in the lesson, it could be associated with cognitive processes. However, when a language learner tries to understand, regulate, and control his or her own strategies, this time that kind of behaviors refers to metacognitive processes.

In addition, metacognition consists of different types of knowledge such as knowing about oneself as a learner, knowing about the task, and knowing about the strategies one can use to carry out the task (Flavell, 1979). These are vital for metacognitive thinking as knowing our strengths and weaknesses help us compensate for our deficiencies and take precautions. In other words, metacognition also makes people assess themselves and change their way of learning accordingly. To illustrate, a person, who knows that he or she gets anxious while giving a speech in front of others in L2, tries to cope with his or her anxiety in different ways, such as tries to regulate breath, to make eye contact or avoid it, and so on. Thus, the individual eventually can give a successful speech.

Furthermore, according to Anderson (2012), metacognition is composed of different components or stages, such as "planning to learn, deciding when to use strategies, monitoring their use, combining strategies where appropriate, and evaluating their effectiveness" (Williams et al., 2015, p. 134). However, these are not consecutive, they may occur at different times or at the same time. Besides, with different perspectives, various kinds of categorizations can be done.

All in all, metacognition has a crucial place in self-regulation and the literature also indicates that the individuals who have good metacognitive abilities are effective learners and facilitating the use of metacognitive strategies leads to more successful language learning and make learners more autonomous and agent. Therefore, teachers have a responsibility to make learners use metacognitive strategies more with various kinds of activities such as keeping diaries, strategy training, and self-assessment checklists. In addition, course book designers also need to take metacognitive strategies into consideration in the process of constructing course materials.

While taking into account all these mentioned so far, the present study aimed to understand the role of goals in metacognition. Thus, the practitioners can make use of goal setting strategies to elevate the learners' metacognitive skills.

1.2. Goals

There are several types of goals which are set by people on their own or by others. Besides, goals can be both short-term and long-term. For instance, people as learners may wish to be successful in a certain lesson or to participate in particular groups to complete a task or to become a successful language teacher in the future, and so on. Thus, to understand goals a great deal of effort was given; however, one of them was the most influential one

which was goal setting theory (Locke & Latham, 1994). According to this theory, there are three essential points which are goals' specificity, their perceived difficulty, and individuals' degree of commitment. Namely, goals should be clear, motivating and achievable. At this point, one should keep in mind that the difficulty should be suitable to the individuals' capacity and they also should include a challenging element so that it can be both motivating and achievable (Williams et al., 2015).

Besides this theory, there is another vital distinction in terms of goals which emphasizes other people's role (Urdan, Ryan, Anderman, & Gheen, 2002). The distinction is succinctly between mastery goals and performance goals (Ames, 1992; Dweck, 1999; Woodrow,2012). Mastery goals could be associated with one's desires to master a new skill or a piece of knowledge. Thus, the individuals who tend to adopt mastery goals may be more motivated to carry out some certain goals than others. Evidently, mastery goals could be associated with positive motional profile (Midgley, 2002). Thus, the teachers should facilitate students to adopt more mastery goals and accordingly she/he can avoid the negative outcomes of testing systems. Moreover, mastery goals can also be separated into two. First one, mastery-approach goals that are directly about one's own prospect, as in the case of an individual who wants to be competent in a foreign language so that he or she can read novels without translation. The other one is mastery-avoidance goals which could be explained by one's desire to master a piece of knowledge again, but the impetus comes from avoiding a negative outcome.

Performance goals, on the other hand, refer to the goals that are set by the individuals who desire to perform to others. Accordingly, self-esteem may play an important role as the individuals who tend to adopt performance goals regulate their actions according to other people's evaluations. Furthermore, these evaluations may be broader than what people assume as they do not only include in positive or negative reinforcements like praising or scolding, but also include in basic feelings like shame, fear, anxiety and so on. (Williams et al.,2015). However, there is also a similar kind of distinction, performance-approach goals which could be associated with "making a good impression before others" and performance-avoidance goals in which the individuals "want to avoid from looking bad before others" (Williams et al., 2015, p. 133). Adopting performance goals which is based on regulating one's acts according to others' desires may give unfavorable outcomes in educational settings. Therefore, the teachers have the responsibility to not to encourage competition in classrooms after certain levels.

In addition, there are several studies that emphasize the role of ethnicity or cultural preferences as they indicated that the individuals coming from Eastern cultures are more likely to adopt performance-avoidance goals than the ones coming from Western cultures (Salili, Chui, & Lai, 2001; Woodrow, 2008).

Lastly, one needs to take into account that these goals can occur separately or simultaneously (Pintrich, 2000; Woodrow, 2012). Besides, the literature indicates a research gap in terms of goals in language classroom settings (Woodrow, 2012).

1.3.The Relationship between Goals and Metacognition

According to previous studies, the relationship between goals and metacognition is an intriguing issue as the studies showed inconsistency. To begin with, the relationship between mastery goals and metacognition was found statistically significant and relatively strong by some scholars (Ames & Archer, 1988; Coutinho, 2007; Dweck & Legett, 1988; Kharazi et al., 2008; Pintrich & DeGroot, 1990; Schmidt & Ford, 2003; Zafarmand et al., 2014). In addition, the studies also found out a link between mastery goals and high self-efficacy, good

metacognitive language learning strategies, and high level of English speaking skills (Woodrow, 2006).

However, the findings on the relationship between performance goals and metacognition differed according to the various studies. Some indicated that there was a positive relationship (Ames & Archer, 1988; Butler, 1993; Coutinho, 2007; Gul & Shehzad, 2012; Kharazi et al., 2008), whereas others indicated that there was no relationship between performance goals and metacognition (e.g., Ford, Smith, Weissbein, Gully, & Salas, 1998).

In addition to these relationships, the relationship between mastery goals and performance goals is important. Even though the previous studies gave negative or unrelated results (Midgley et al., 1998), the more recent ones generally indicted positive results (Barron & Harackiewicz, 2001; Pintrich, 2000; Woodrow, 2012).

In summary, the studies which include academic achievement indicated that metacognition is a strong predictor (Dunning, Johnson, Ehrlinger, & Kruger, 2003). In addition, setting goals both mastery and performance goals facilitated the academic achievement (Roebken, 2007). Both metacognition and goals play a crucial role in academic success; therefore, the present study aimed to understand their relationship. To this end, the following research questions were formulated to guide the present study.

1) What are the participants' perceived levels of metacognitive awareness, mastery and performance goals?

2) Is there any statistically significant relationship among metacognition, mastery goals and performance goals?

3) Is there any change in the relationship between mastery goals and metacognition when the performance goals are in control?

2. Methodology

2.1. Research Design

The quantitative research design was adopted for the present study. Two inventories were utilized to understand the participants' goal orientation and metacognitive awareness along with the demographic information.

2.2. Setting and Participants

For the present study, a total of 118 participants were selected with the help of convenient sampling technique from the EFL students who were studying at a major state university in Turkey. The participants ranged in age from 18 to 22 ($M = 19.30$ years, $SD = .85$). In addition, Table 1 shows the gender distribution of the participants.

Table 1. *Gender distribution of the participants*

Variables	Gender	N	Mean	SD	%
Gender	Male	28	-	-	23.7
	Female	90			76.3

2.3. Instrumentation

2.3.1. Metacognitive Awareness Inventory

The Metacognitive Awareness Inventory (Shraw & Dennison, 1994) was used to identify the participants' perceived levels of metacognitive awareness. The inventory consisted of 52

statements on a 5 point (from 1= strongly disagree to 5= strongly agree) Likert scale. The internal consistency of the scale was α=.93 in the present study.

2.3.2. Goals Inventory

To measure the participants' mastery goals and performance goals, Goals Inventory in a 7-point Likert scale format (from 1= strongly disagree to 7 = strongly agree) was conducted (Roedel, Schraw, & Plake, 1994). The inventory consisted of two sub-categories; mastery goals were measured with 12 items and performance goals were measured with 5 items. The internal consistency of the subscales is α=.84 and α=.84, respectively.

2.4. Data Collection Procedure and Analysis

The current study was carried out with English majors from a state university at the end of the fall semester of 2017- 2018 academic year. The inventories were distributed to the participants and they completed the inventories in 20 minutes for each. After gathering the data, descriptive statistics such as frequencies, percentages, and mean scores were calculated with the help of IBM SPSS 21. After the normality check, to understand if there is any statistically significant relationship among metacognition, mastery goal, and performance goal, Pearson Correlation Coefficients were utilized. In addition, to further understand the relationship between metacognition and mastery goals a Partial Correlation was conducted.

3. Results

To address the first research question, descriptive statistics for metacognitive awareness ($M = 3.79$, $SD = .45$), mastery goals ($M = 4.49$, $SD = 1.0$) and performance goals ($M = 4.47$, $SD = 1.4$) were utilized. In addition, results showed that 48.3% of the participants had high, 28% of the participants had mid, and 23.7% of the participants had low metacognitive awareness (Table 2).

Table 2. *Descriptive Statistics for metacognition, mastery goals, and performance goals*

	N	Min	Max	M	SD
Metacognition	118	2.79	4.88	3.7903	.45368
Mastery Goals	118	2.25	6.58	4.4915	1.00228
Performance Goals	118	1.00	7.00	4.4746	1.44478
Valid N (listwise)	118				

For the second research question, a Pearson correlation was conducted to understand the relationship among metacognition ($M=3.79$, $SD = .47$), mastery goals ($M = 4.49$, $SD = 1.0$), and performance goals ($M = 4.47$, $SD = 1.4$) (Table 3).

Table 3. *Correlations among metacognition, mastery goals, and performance goals*

Variables	1	2	3
1. Metacognition	↓		
2. Performance Goals	.145	↓	
3. Mastery Goals	.565[**]	.291[**]	↓

**. Correlation is significant at the 0.01 level (2-tailed).

The Pearson correlation coefficients indicated that there is a statistically significant relationship between metacognition and mastery goals ($r(118) = .565$, $p =. 000$) which

indicates a large correlation (Cohen, 1988), and 31 per cent shared variance. However, the relationship between metacognition and performance goals is statistically non-significant (r(118) = .145, p > .005). In addition, the results also indicated a statistically significant relationship between mastery and performance goals (r(118) = .291, p = .001).

To address the third research question, whether there is any change in the relationship between metacognition and mastery goals when the performance goals are stabilized, a partial correlation test was employed to make further judgments on the results (Table 4).

Table 4. *Partial Correlation Results*

Control Variables		1	2
	1. Metacognition	↓	
Performance Goals	2. Mastery Goals	.552[**]	↓

The partial correlations indicated that the relationship between metacognition and mastery goals (r (115) =.552, p=.000) is still significant even though it loses its strength in small amounts.

4. Discussion

The main purpose of the study was to investigate the possible relationship between goals and metacognitive awareness. The results indicated that there was a significant relationship between mastery goals and metacognitive awareness. In addition, these results of the Pearson correlation and partial correlation are compatible with the existing literature in terms of the relationship between metacognition and mastery goals (Ames & Archer, 1988; Coutinho, 2007; Dweck & Legett, 1988; Kharazi et al., 2008; Pintrich & DeGroot, 1990; Schmidt & Ford, 2003; Woodrow, 2006; Zafarmand et al., 2014).

However, the relationship between metacognition and performance goals was statistically non-significant in the present study. In terms of this relationship, the previous studies also indicated different results. Some indicated that positive relationship (Ames & Archer, 1988; Butler, 1993; Coutinho, 2007; Gul & Shehzad, 2012; Kharazi et al., 2008), while others found no relationship between these constructs (Ford et al., 1998).

In addition, the results also revealed that there was a significant relationship between mastery goals and performance goals which means that different goal orientation can occur at the same time. This is also in harmony with the existing literature (Barron & Harackiewicz, 2001; Pintrich, 2000; Woodrow, 2012).

The results also indicated the participants' perceived levels of mastery goals, performance goals and metacognitive awareness. As it can be seen in Table 2, the participant perceived levels of mastery goal orientation and performance goal orientation was approximately at the same level, this can be explained by cultural preferences as the previous studies indicated that Eastern cultures more tended to adopt performance goals than Western cultures. However, the sample group may be classified as Middle Eastern (Salili et al., 2001; Woodrow, 2008).

However, seldom cultural preferences cannot explain such kind of result. One needs to consider the context and the education system in it. It is a known fact that in foreign language context the individuals do not have many opportunities to contact with the target community and this facilitates performance goals more by emphasizing competition and testing systems (Williams et al., 2015).

5. Conclusion

Metacognition is a strong predictor of academic achievement (Dunning et al., 2003; Öz, 2014), suggesting that learners' metacognitive awareness should be promoted to enable them to be more successful. The findings of the present study indicated statistically significant correlation between mastery goals and metacognition. Therefore, engaging students in activities on goal setting, especially enabling students to set mastery goals for themselves, may help them be more successful. This result also can be interpreted in relation to context which is EFL for the present study. In EFL contexts the individuals do not have enough opportunities to interact with the target community. Thus, the education system is competitive, and testing based, and this system makes learners adopt more performance goals than mastery goals (Anderman, Patrick, Hruda, & Linnenbrink, 2002). However, adopting performance goals is not a desired outcome of the learning process. To avoid this, the teachers may contribute well even though there may be oppositions from their authorities or even the societal values (Williams et al., 2015).

Furthermore, it should be kept in mind that languages are taught as a compulsory course in EFL contexts. Thus, the learners may not have a goal or do not have an idea why they have to learn that language. Therefore, an investigation on their goal orientation should be carried out repeatedly as their orientation could change within time (Dörnyei, 2001).

Even though the relationship between performance goals and metacognition was not statistically significant in the present study, the relationship between mastery goals and performance goals was statistically significant. Thus, the role of performance goals in educational settings should not be denied. In addition, according to the previous studies, setting goals – no matter mastery or performance goals – facilitated academic achievement (Roebken, 2007). Therefore, goal setting has a crucial place in educational settings.

Accordingly, to elevate goal setting, the activities and materials should not conflict with the learners' interests and preferences such as younger learners may prefer more technology based or online tools more than adults; therefore, a need analysis may be required.

For setting long-term goals identification and discussion type of activities could be carried out while for setting short-term goals certain goals can be supported such as "how many words they will learn this week; how many books in the target language they will read; how they will practice speaking the language; how many television programs they will watch in the language" (Williams et al., 2015, p. 103).

However, like all studies this one also has limitations especially in terms of only using quantitative methodology. This gives us a drawback by looking into only the perceived levels of the participants' goal setting and metacognitive awareness. Thus, larger sample studies may be more insightful. Furthermore, the gender distribution is important as the present study did not have equivalent gender distribution due to the context as in Turkey females tend to choose foreign language fields more than males.

In conclusion, the present study indicated that goal setting has a vital role in metacognitive awareness and it is an essential part of self-regulated language learning. Thus, to make students self-regulatory learners, language teachers, curriculum designers, and authorities may try to elevate the students' metacognitive awareness and make them set goals so that the students may become autonomous and good agents, and eventually, successful language learning can occur.

References

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261-271

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivational processes. *Journal of Applied Psychology, 76,* 478-487.

Anderman, L. H., Patrick, H., Hruda, L. Z., & Linnenbrink, E. A. (2002). Observing classroom goal structures to clarify and expand goal theory. In C. Midgley (Ed.), *Goals, goal structures and patterns of adaptive learning* (pp. 243–278). Mahwah: Lawrence Erlbaum.

Anderson, N. J. (2012). Metacognition: Awareness of language learning. In S. Mercer, S. Ryan & M. Williams (Eds.), *Psychology for language learning: Insights from research, theory and practice* (pp. 169-187). London: Palgrave Macmillan UK.

Barron, K. E., & Harackiewicz, J. M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology, 80*(5), 706–722.

Butler, R. (1993). Effects of task-and ego-achievement goals on information seeking during task engagement. *Journal of Personality and Social Psychology, 65*, 18-31.

Coutinho, S. (2007). The relationship between goals, metacognition, and academic success. *Educate, 7* (1), 39-47.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12* (3), 83-87.

Dö rnyei, Z. (2001). *Motivational strategies in the language classroom.* Cambridge: Cambridge University Press.

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development.* Hove: Psychology Press.

Dweck, C. S., & Leggett, E. S. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95,* 256-273.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of psychological inquiry. *American Psychologist, 34,* 906-911.

Ford, J. K., Smith, E. M., Weissbein, D.A., Gully, S. M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology, 83,* 218-233.

Gul, F., & Shehzad, S. (2012). Relationship between metacognition, goal orientation and academic achievement. *Procedia-Social and Behavioral Sciences, 47,* 1864-1868.

Kharazi, S. A. N., Ezhehei, J., Ghazi Tabatabaei, M., & Kareshki, H. (2008). An investigation of the relationships between achievement goals, self-efficacy and metacognitive strategies: Testing a causal model. *Journal of Psychology and Education, 38,* 69-87.

Locke, E. A., & Latham, G. P. (1994). Goal setting theory. In H. F. O'Neil, & M. Drillings (Eds.), *Motivation: Theory and research* (pp.13-29). Hillsdate, NJ: Lawrence Erlbaum Associates.

Midgley, C. (Ed.). (2002). *Goals, goal structures and patterns of adaptive learning.* Mahwah: Lawrence Erlbaum.

Midgley, C., Kaplan, A., Middleton, M., Maehr, M., Urdan, T., Anderman, L., et al. (1998). The development and validation of scales assessing students' achievement goal orientations. *Contemporary Educational Psychology, 23*(2), 113–131.

Öz, H. (2014). The relationship between metacognitive awareness and academic achievement among English as a Foreign Language teachers. In J. Huang & A. C. Fernandes (Eds.) *Non-native language teaching and learning: Putting the puzzle together* (pp.139-167). New York, NY: Untested Ideas Research Center.

Öz, H. (2015). Investigating pre-service English teachers' metacognitive awareness. In H. Öz (Ed.), *Language and communication research around the globe: Exploring untested ideas* (pp. 35-58). New York, NY: Untested Ideas Research Center.

Pintrich, P. R. (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology, 92*(3), 544–555.

Pintrich, P. R., & DeGroot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82,* 33-40.

Roebken, H. (2007). The influence of goal orientation on student satisfaction, academic engagement and achievement. *Electronic Journal of Research in Educational Psychology, 5*(3), 679-704.

Roedel, T. D., Schraw, G., & Plake, B. S. (1994) Validation of a measure of learning and performance goal orientation. *Educational and Psychological Measurement 54*(4), 1013-1021.

Salili, F., Chiu, S. Y., & Lai, S. (2001). The influence of culture and context on stu- dents' motivational orientation and performance. In F. Salili, C. Y. Chiu, & Y. Y. Hong (Eds.), *Student motivation: The culture and context of learning.* New York: Plenum Publishers.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4), 460-475.

Schmidt, A. M., & Ford, K. (2003). Learning within a learner control training environment: The interactive effects of goal orientation and metacognitive instruction on learning outcomes. *Personnel Psychology, 56,* 405-429.

Urdan, T., Ryan, R. M., Anderman, E., & Gheen, M. H. (2002). Goals, goal structures and avoidance behaviours. In C. Midgley (Ed.), *Goals, goal structures and patterns of adaptive learning* (pp. 55–84). Mahwah: Lawrence Erlbaum.

Williams, M., Mercer, S., & Ryan, S. (2015). *Exploring psychology in language learning and teaching.* Oxford: Oxford University Press.

Woodrow, L. J. (2006). A model of adaptive language learning. *The Modern Language Journal, 90*(3), 297–319.

Woodrow, L. J. (2008). *Adaptive second language learning: The case of EAP students.* Saarbrucken: VDM Verlag Dr Muller.

Woodrow, L. (2012). Goal orientations: Three perspectives on motivation goal orientations. In S. Mercer, S. Ryan, & M. Williams (Eds.), *Psychology for language learning: Insights from theory, research and practice* (pp. 188-202). Basingstoke: Palgrave Macmillian

Zafarmand, A., Ghanizadeh, A., & Akbari, O. (2014). A structural equation modeling of EFL learners' goal orientation, metacognitive awareness, and self-efficacy. *Advances in Language and Literary Studies, 5*(6), 112-124.

EDUCATIONAL ANALYTICS ON AN OPENCOURSEWARE

Seher Balbay [iD]

Middle East Technical University

seherb@metu.edu.tr (correspondence)

Selcan Kilis [iD]

Giresun University

k1selcan@gmail.com

Seher Balbay has been an instructor at Middle East Technical University since 1997. She received her MA from the same university in 2000. She's working towards her PhD degree in ELT at Hacettepe University.

Selcan Kilis is an assistant professor in the Division of Curriculum and Instruction in Giresun University. She received her PhD degree from the Department of Computer Education and Instructional Technology at METU in 2016.

EDUCATIONAL ANALYTICS ON AN OPENCOURSEWARE

Seher Balbay
seherb@metu.edu.tr

Selcan Kilis
k1selcan@gmail.com

Abstract

Analytics as one of the recent fields in technology-based learning offers many benefits to educators, instructors, and administrators to improve the efficiency and quality of alternative educational materials, and learning experience through tracking and storing students' log data on web platforms over an extended period of time. This mixed-method study investigates students' log data retrieved from the opencourseware (OCW) specifically launched for a required academic English speaking skills course offered at Middle East Technical University in Turkey with the aim of enhancing the quality and efficiency of the materials available for the course. By understanding the reasons behind students' behaviors via the interviews conducted with 50 students on this online courseware, this study also aims to provide useful practical hints to the instructors and guide them to act on future decisions. The analyzed data is based on learner behavior with a specific emphasis on average view duration, likes and dislikes, and comments. This study can serve as a starting point to guide and provide the instructors and administrators about the future of the aforementioned course which is also offered in a rotational hybrid learning format where the effectiveness of online materials gain even more importance.

Keywords: learning analytics, online video-based learning, log data, academic speaking skills, opencourseware

1. Introduction

1.1. Why Track Learning Analytics?

One of the major developments in data collection on educational statistics is the availability of 'big data' using the data visualization techniques online course platforms lets educators access to. The interest in this robust data draws attention to tracking the activities of students providing teachers, especially those in decision-making administrative positions with 'navigation or behavior-focused statistics' (Bull & Kay, 2016, p. 311). Although, publications on the interpretation of such data is unprecedented in English language teaching programs, it is directive value is undeniable. That is the learner analytics of log data can be guiding for future online tools and materials in similar teaching settings.

A related field is educational data mining. EDM is a more generic term which encompasses both LA and Academic analytics. Although the investigators, methods and findings of Learning Analytics (LA) and Educational Data Mining (EDM) overlap to a great extent some researchers differentiate between the two by claiming that data mining encompasses both learning analytics and academic analytics. Academic analytics is basically defined as 'the process of evaluating and analyzing organizational data received from university systems for reporting and decision making reasons' (Campbell & Oblinger, 2007).

'Learning Analytics' (LA) is the 'measurement, collection, analysis and reporting of such big data about learners' behaviors with the intention of understanding and optimizing learning and the environments in which it occurs' (Sclatter, Peasegood & Mullan, 2018).

LA is a new area of interest that bears a crucial importance in the era of technology in education. Simply put, learning analytics traces the learning process-related online data and reveals systematic measurement of the frequencies the online educational tools are used. The learner analytics field is data-driven and atheoretical. Online learning management systems today are becoming common in all levels and fields of education today, and the data retrieved from their use can be enlightening about which sort of materials are on demand by learners. This quantitative data available can shape the future of blended learning platforms, as it can be a predictor for detecting student preferences of course-related resources. It must be made clear that this statistical data does not predict the achievement of learning outcomes, but has the potential to impact the success (Arnold & Pistilli, 2012; Barber & Sharkey 2012; Gibson & de Freitas 2015; Mah, 2016). Yet, it is a valuable source to receive feedback on the learning process that takes place on the internet. Bienkowski, Feng and Means (2012) state that learning analytics provides institutions with ample opportunities to support the student learning process, and to enable personalized learning. That is; with the spread of online learning tools, personalized learning options will be developed based on the reports of learner preferences of online learning tools and materials.

This easy to retrieve data is also easy to interpret thanks to the skillfully designed visualizations of analytics engines. Hence, it makes it possible for educational institutions to use the experience of the past and plan the future investments on education. Although many educational institutions are not ready to exploit learner analytics today, it is an undeniable fact that the future of online learning tools depend on it.

Learner analytics use data generated from learner activities which can sometimes be not only watch time, but simply clicks, participation in online forums, or computer assisted testing (Tempelaar, Rienties, Giesbers, 2014). Shum and Crick (2012, p.3) state that learning activity generated data is also proof of an overview of 'students' values and attitudes which are fed back to students and teachers through visual analytics'. Data extracted from online institutional learning management systems by measuring learner preferences can also be used for motivation and engagement-related research in an era of intense online learners when making plans concerning longitudinal learning infrastructure.

By the same token, Crede and Niehorster (2012) suggest that learner analytics data can be studied as an indicator of academic performance emotional and social factors. It is of significance at this point to iterate that learner analytics data alone would not be a direct and explicit proof of any correlations without a demographic overview of the participants in a study. At this point, the background information on the context of studies based on learner analytics bears undeniable importance. Knowledge of course design, instructor intentions and student, institution and the design of the Moodle, Web 2.0 tool or the open courseware that the data is extracted from is of utmost importance when analyzing learner analytics data since these are the major factors that determine 'which variables can meaningfully represent student endeavor and engagement' in activities provided online (Macfadyen & Dawson, 2010, p. 597).

The goals of learner analytics are 'predicting learner performance, suggesting new learner resources, increasing reflection and awareness, enhancing social learning environments and detecting undesirable learner behaviors' as listed by Verbert, Manauselis, Drachsler and Duval (2012, p.138). Learner analytics provides self-evident data for online educational material which is less preferred by students, which can enlighten material designers as to the

type of content and means to opt for knowing which components of sources best serve the targeted population and objectives. Hence, bearing in mind the directive feature of the objective data learner analytics provides, it is apparent that this new field deserves further and longitudinal research.

1.2. Learning Analytics, the Research Field of 21st Century Classrooms

Learning analytics literature has recently been published citing the many benefits of both making course sources and interaction online through a variety of platforms and also by referring to making the input material for courses accessible online.

It has been observed that students use social networks extensively as a learning tool (Agudo-Peregrina, Iglesias-Pradas, Conde-Gonzales, Hernandez-Garcia, 2014). One study observing the effects of social networks on performance was conducted on 300 medicine faculty students. The study concluded that the use of social networks has a predictive influence on the academic performance of the students. Another study conducted by Rienties, Hernandez Nanclares, Hommes and Veermans (2014) found out that 30-80% of learning occurred outside formal settings of education. Personal preferences of using online course material bears importance.

Researchers have been vocal in expressing how individual learning styles are considered when preparing sources that students can access in online course material sharing platforms studying the handouts, videos or slides at their own pace. Personalization of teaching is one of the most significant benefits of interfaces that present course material (Bull & Kay, 2016). Online learning management systems prompt self-regulated learning and metacognitive skills too since such tools can on particular occasions play the role of the teacher.

School leaders and policy makers are also stakeholders of learner analytics, which is emphasized by Long and Siemens (2011). Actually, the policy makers who allocate the personnel and funding of educational institutions need to collect and understand the learner analytics data to be able to make sound decisions about the efficiency of their future plans and investments. Institutional use of such 'big data' is the starting point of learner analytics. The increasing interest in collecting and interpreting this big data provided by learner analytics will keep the educational and governmental institutions, such as the university decision making boards, administrators, the higher education councils, and the Ministry of National Education as its beneficiaries. Apparently, 'Educational Data Mining' is expected to lend itself to be used for policy making at all levels.

Another benefit of online course input is the undeniable fact that it is cost-effective (Bull & Kay, 2016). Cloud systems are widely popular today, yet opencourseware platforms of educational institutions present an alternative to the limited or costly expanses of cloud systems. Long term store is also possible thanks to these opencourseware interfaces.

Bull and Kay (2016) emphasize the fact that learning technology has been exponentially pervasive in the past few years and hence shifted the perception of the means of learning in general making use of online resources outside the time and location constraints of the traditional classrooms in their research on SMILI☺ (Student Models that Invite the Learner In), an open learner framework created to provide a framework of other open learner models.

Although this research paper focuses on the frequencies in the use of the Middle East Technical University open courseware for the Academic Oral Presentation Skills course, which is an evident sign of the change in the educational tools that we use, the current study still refers to the nature of learning tools today directly, since the aim of this study is to analyze the frequency of the use of online course materials and interpret this frequency

analysis via the interviews conducted with students. Therefore, being one of the pioneers of such learning analytics studies in its field, it is going to be a forerunner of the use of effective computer-based or computer-dependent teaching systems in the English Language Teaching field. The analysis of the popularity of the sources and investigating the reasons behind this popularity will also be revealing in determining the student choices, hence will shed light on the type of future materials. The study intends to make sense of student preferences. Drawing on the statistical analysis of student behaviors, the broad knowledge LA provides on the use of different course materials will be data-driven.

Bearing the above benefits of LA in mind, this research study explores the answers to the following research questions (RQ):

RQ1: Which type of supplementary online course materials were used most by the students taking the Oral Presentation Skills Course?

RQ2: What are students' reasons for preferring certain types of materials over others?

2. Method

This mixed-method study aims to investigate which type of supplementary online course materials were used most by the students on the OCW platform for the Oral Presentation Skills Course in addition to determine students' reasons for preferring certain types of materials over others. With this aim, the current research analyzed the frequency of clicks on the materials made available to students on the open courseware of the required Academic Oral Presentation Skills course offered to an average of 1000 students each semester.

2.1. Research Context

The research context of the study is the Middle East Technical University, one of the most prominent English-medium instruction universities of Turkey where the Academic Oral Presentation Skills course is a required freshman English course to students from all departments. The course is offered every semester including the summer school. During the fall and spring semesters, about 1000 students are offered the course. The number of students that can take the course during the summer school is about half that number, i.e. 500 students on average. In the course of the time that the open courseware was operational, an estimate of 5000 students were offered the course. While some instructors teaching the course referred their students to the open courseware, some instructors preferred to refer their students to their own supplementary materials. Still, as the name suggests, the 'open' courseware was available to all students.

2.2. Participant (Subject) Characteristics

The blueprints of an average of 1000 students during five semesters (a total of 5000) students' on the OCW platform were collected. The students taking the Academic Oral Presentation Skills Course who made use of the materials on the opencourseware were from different faculties (Faculty of Architecture, Arts and Sciences, Economic and Administrative Sciences, Education, and Engineering). The course is offered to all students who have passed the two prerequisite Academic English courses offered in their first year at their departments. So students were from 2nd to 4th year students, and their ages therefore ranged from 21 to 25 mostly.

With regard to the qualitative part, the most clicked sources were identified and 50 students were chosen according to convenience sampling method to conduct interviews with, with the aim of investigating the reasons behind the popularity of the particular sources that stand out in the numerical learning analysis.

The students studying at the Middle East Technical university are digital technology natives, very efficiently making the best use of the fastest internet in Turkey (10 Gigabit Ethernet technology), available at all locations on the university campus ("Campus Backbone Network METU-NET | Computer Center", n.d.). The campus is well-equipped with computer laboratories at both department buildings and dormitories. Smartphones are an indispensable part of these students' lives. A former unpublished research, Blended Learning in a Speaking Skills Course reports that 99% of students have smartphones and are very used to using them for both educational and entertainment purposes in addition to communication (Balbay & Kilis, 2017).

2.3. Research Design

This mixed-method study applied explanatory sequential design which is also called two-phase model (Cresswell & Plano Clark, 2011). Its rationale is that findings retrieved from the quantitative data provide a general picture of the research problem, and then the qualitative data is analyzed to refine, extend, or explain the general picture. Therefore, this research design type basically consists of first collecting quantitative data to get a general picture of the problem investigated, and then collecting qualitative data to refine the results from the quantitative data (Cresswell, 2012). The process of explanatory research design is presented in Figure 1.
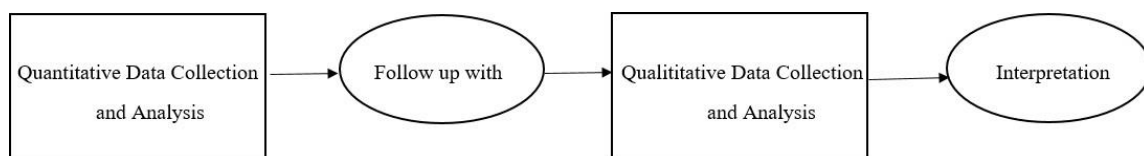

Figure 1. Explanatory sequential design (adapted from Cresswell, 2012)

Quantitative data is used for the first research question whereas qualitative data is for the second research question.

2.4. Data Collection and Instruments

Quantitative data is composed of students' log data on the open courseware platform used in the Academic Oral Presentation Skills course. The open courseware was launched in the fall semester of 2016 and the period that the analysis covers is the fall, spring and summer semesters of the academic year 2016-2017, and the fall and spring semesters of 2017-2018 academic year. Students' log data was used to determine which type of supplementary online course materials were used most by the students taking the Oral Presentation Skills Course. Moreover, qualitative data was collected through a semi-structured interview protocol to learn about students' reasons for preferring certain types of materials over others. The semi-structured interview questions were simply 'Which online sources did you benefit most from and why?'.

2.5. Data Analysis

Quantitative data was analyzed using descriptive statistics. Qualitative data was analyzed with deductive content analysis. Interview data was categorized under the most common reasons and coded by the two researchers of this study for inter-rater reliability. The coding agreement by two raters was found to be at about 90% percentage.

3. Results

The first research question 'Which type of supplementary online course materials were used most by the students taking the Oral Presentation Skills Course?' was analyzed with students' log data on the OCW platform. First of all, course material in each unit on the OCW platform were analyzed by descriptive statistics. Table 1 and figure 2, below display the distribution of the use of supplementary course material according to the units of the course book.

Table 1. *Descriptive statistics about course units on the OCW*

Units	Frequency	Percentage
Welcome	4083	7.98
Unit 1	8834	17.28
Unit 2	11417	22.33
Unit 3	13127	25.67
Unit 4	11735	22.95
Other	1939	3.79

Total logs = 51135

According to Table 1, students clicked mostly the materials on Unit 3 (n=13127, ~26%), followed by Unit 4 (n=11735, ~23%) and Unit 2 (n=11417, ~22%), and then Unit 1 (n=8834, ~17%). They also visited the 'Welcome' section on the OCW at about 8% (n=4083). Lastly, other logs of a total of 51135 logs were at about 4% percent (n=1939) as seen in Figure 1. Overall, the findings indicated that students mostly visited each unit almost about the same frequency.
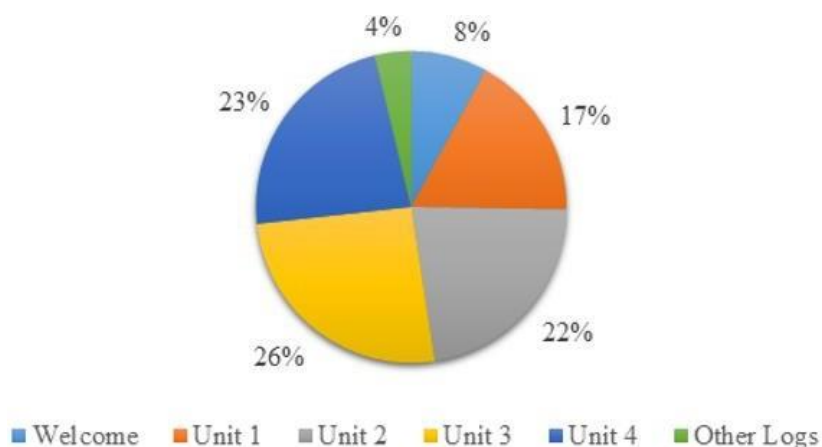

Figure 2. Distribution of students' blueprints based on course units

Findings about document types that students clicked on the OCW platform are presented in Table 2.

Table 2. *Descriptive statistics about the types of supplementary materials on the OCW*

Type of Document	Frequency	Percentage
File	36139	70.67
Url	14802	28.95
Forum	194	.38

Total logs = 51135

Students mostly used file type document (n=36139, ~71%) on the OCW platform. they clicked and benefitted from URL type document at the ratio of one third of file type document. They barely visited forum on the OCW platform. Findings about types of course content students clicked on the OCW platform with descriptive are presented in Table 3.

Table 3. *Descriptive statistics about type of course content on the OCW*

Content	Frequency	Percentage
Custom-made Input Videos for 211	3943	7.71
Course-related Word Files (Guidelines & Rubrics)	15755	30.81
Sample Presentations	2461	4.81
Practice Materials	4960	9.70
Listening Practice Materials	5461	10.68
Relevant YouTube Videos	1623	3.17
Input Slides	10784	21.09
Games	1562	3.05
Other Logs	4586	8.97
Total Logs	51135	100.00

As can be seen in Table 3, students used mostly course-related word files such as guidelines and rubrics (n=15755, ~31%), followed by input slides (n=10784, ~21%). They benefited from practice materials (n=4960, ~10%) at a ratio of one-tenth of a total log of 51135 logs. After practice materials, the highest percentage belongs to the other logs at a ratio of about nine percent of total logs. Students benefited from custom-made input videos for the course that the current study was conducted at a ratio of about eight percent of total logs. Their logs ratio for sample presentations was about five percent (n=2461, ~5%), relevant YouTube videos was (n=1623, 3%), and lastly games was (n=1562, 3%). Finally, findings about the most and the least clicked documents on the OCW platform with descriptive are presented in Table 4.

Table 4. *The most and the least clicked documents or URL*

The Most and The Least Clicked Documents or URL	Frequency of Clicks
The Five Highest Clicked Documents or URL	
Marketing Presentation File	1555
Science And Technology Presentation File	1463
Course Outline File	1278
Weekly Schedule File	1212
Mini Presentation Rubric File	1003
The Five Least Clicked Documents or URL	
Final Presentations Playlist URL	62
Listening Material - Eternal Sunshine File	88
Video Sources - Your Food Is Shrinking URL	121

Video Sources - Life After Death By PowerPoint URL	133
Rubric - Debate Jury-sheet File	150

As can be seen in Table 4, students mostly clicked and used the file Marketing Presentation file n=1555), followed by Science and Technology Presentation file (n=1463), course outline file (n=1278), weekly schedule file (n=1212), and then Mini Presentation Rubric file (n=1003). On the other hand, the five least clicked documents or URL that students used or clicked were Rubric - Debate Jury-sheet file (n=150), Video Sources - Life After Death By PowerPoint URL (n=133), Video Sources - Your Food Is Shrinking URL (n=121), Listening Material - Eternal Sunshine File (n=88), and finally Final Presentations Playlist URL (n=62).

The second research question 'What are students' reasons for preferring certain types of materials over the others?' was analyzed with qualitative data obtained through semi-structured interviews with 50 students. The two researchers coded the recurring themes for students' reasons for preferring certain type of materials. The coding agreement by two raters was found to be at about 90% percentage.

To start with, when students were asked why they referred to the OCW most for Unit 3 materials, they repeatedly stated that exercises on visual representation of numerical data were to be asked in the exam and they referred to those materials for exam preparation practice. Secondly, most students said that the first major presentation they gave was in unit 3, so they needed the guidelines, slides and examples on the OCW for the Marketing Presentation. Apparently, in the two most recurring answers collected during the interviews assessment plays a major role. When asked why they though the first section on the OCW was not clicked frequently, students revealed that the outline and the weekly schedule posted on the Welcome session were two main handouts that they never needed to refer to throughout the semester. The almost equal distribution of the clicks among the units were supported with an answer from the students that focuses on assessment again. Students repeatedly revealed that they checked the guidelines, informative videos and slides for the tasks that were going to be evaluated and graded and since there are graded tasks in each unit, they checked the materials for each unit almost on an equal frequency.

A majority of the students reported that they preferred handouts to slides and videos because checking information on handouts is what they are familiar with. Handouts also come in handy, they said. Some students stated that they can take a screenshot of the handouts on their phones to refer to them again when they need to. 'The Urls for videos were fun to watch but not practical when we are doing self-study' one student claimed.

When students missed class they went back to what was covered during that session, they said during the interviews. That is why the percentage of clicks of the slides available on the OCW is among the highest. Students, without any exception stated that they needed to practice listening because they did not feel confident about their listening skills. Hence, the listening practice materials were the most clicked materials on the OCW.

Although numerical data does not support the argument that the course-related games made available on the OCW were popular, students preferred to refer to them a lot during the interviews. The games on the OCW were mostly group games that the whole class played together. That is why the individual click rate is not high. However, the majority of the students mentioned them when they were asked which materials on the OCW they used most and why. They said that being an anonymous player in especially the Kahoot games, relieved the stress on their shoulders and the students who played in pairs were even more relieved

because they had the chance to ask for their peer's approval before they hit one of the multiple choice alternatives on their phones.

3. Discussion and Conclusion

In this empirical study into student behavior different categories of data sources were examined to generate a knowledge of student preferences of online material available on the Middle East Technical University opencourseware. The data collected was from the last two years. It comprises of an input of student behavior from a relatively long period of time during which the Academic Oral Presentation Skills course was offered. During the course of time studied, the course was offered for five semesters including the summer school to an average of five thousand students. Due to this long course of observation period, the data represents a semi-static period of the use of supplementary online course material, it therefore dynamic in nature.

This paper has reflected on the use of the open courseware materials of the Academic Oral Presentation Skills course offered as a required course to students from all the departments at the university. The study was based on objective, numerical data based on clicks of students, later discussed in group interviews with students. Although the qualitative data gathered during discussions provides reasons for students to prefer to 'click' certain materials provided online, the clicks themselves may not be a sign of to make assumptions on the reasons behind them, that is to say, as Macfayden and Dawson (2010, p 597) put it 'simple clicking behavior in a learning management system is at best a poor proxy for actual user behavior of students'. While an account of the clicks of students cannot be considered as a sign of their learning, simple clicks alone can be attributed to intention and motivation to study, practice and explore course-related material, hence, can still be considered as a salient proof of interest, which embraces the growth of even more technology integration in education.

First generation technology-driven learning analytics provides any stakeholder involved in the education process, including teachers, instructors, policy makers, and curriculum and material designers, educational Web 2.0 tools designers with insights from data-driven educational research as to the most efficient online learning tools.

Shum and Crick (2012, p. 2) dwell on the invaluable consequences of the collection and interpretation of the growing evidence provided by learner analytics within educational research. It is claimed that 'learners' orientation towards learning their own learning dispositions significantly influence the nature of their engagement with new learning opportunities'.

There are some potential limitations of this study. Firstly, there are limitations to the generalizability of the findings. The more and detailed analytics of users' behaviors on online platform can be recorded and gathered for further and detailed analytics and explanation on their usage behaviors. Secondly, this study focused merely on users' behaviors on the OCW platform. However, some guest users on the course page due to being an opencourseware page might affected the results. Finally, it would be better for learning analytics demonstrate change over time (Goggins, Galyens, Petakovic, Laffey, 2016), however the subjects of this current study during five semesters were different.

Future research on the use of open courseware may make use of eye-trackers that provide data on focus and gaze (Kardan & Conati, 2012). Even more detailed inquiry is possible via emotion sensors which are promising in the education-related research (Arroyo, Cooper, Burleson, Woolf Muldner, Christopherson, 2009). There is no doubt that as technology becomes even more pervasive in education, the reflections of data-driven student behavior will gain even more importance in predicting the future of educational materials and

preparing our mindset as educators accordingly to meet the growing demand of online learning contexts.

References

Agudo-Peregrina, A. F., Iglesias-Padas, S., Conde-Gonzalez, M. A., & Hernandez-Garcia, A. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior, 31,* 542-550. http://dx.doi.org/10.1016/j.chb.2013.05.031.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington, DC: American Psychological Association.

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In *LAK '12 Proceedings of the 2nd international conference on learning analytics and knowledge*. New York: ACM.

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. *AIED,* 17-24.

Balbay, S., Kilis, S. (2017). Blended learning in a speaking skills course. In the proceedings of *3rd International Congress on Education, Distance Education and Educational Technology (ICDET)*, 24-25 November 2017, Antalya: Turkey.

Barber, R., & Sharkey, M. (2012). Course correction: Using analytics to predict course success. In LAK '12 Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 259–262). New York: ACM.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. *US Department of Education, Office of Educational Technology* (pp. 1-57).

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Journal of Educational Technology & Society, 15*(3). http://dx.doi.org/10.1145/2330601.2330616.

Bull, S., & Kay, J. (2016). SMILI☺: A framework for interfaces to learning data in Open Learner Models, learning analytics and related fields. *Artificial Intelligence in Education, 26*(1), 293-331.

Campbell, J.P., Oblinger, D.G. (2007). Academic Analytics, Educause.

Campus Backbone Network METU-NET | Computer Center. Retrieved 22 May 2018, from https://bidb.metu.edu.tr/en/campus-backbone-network-metu-net

Chapelle, C., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge, UK: Cambridge University Press.

Crede, M., & Niehorster, S. (2012). Adjustment to college as measured by the student adaptation to college questionnaire: A quantitative review of its structure and relationships with correlates and consequences. *Educational Psychology Review, 24*(1), 133-165. http://dx.doi.org/10.1007/s10648-011-9185-5.

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson Education, Inc.

Creswell, J. W. & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. SAGE Publications, Inc.

Dörnyei, Z. (1998). Motivation in second and foreign language learning. Language Teaching, 31(3), 117-135. http://dx.doi.org/10.1017/S026144480001315X

Dörnyei, Z. (2007). *Research methods in applied linguistics: Quantitative, qualitative, and mixed methodologies*. Oxford, UK: Oxford University Press.

Dörnyei, Z., & Csizér, K. (2002). Some dynamics of language attitudes and motivation: Results of a longitudinal nationwide survey. *Applied Linguistics, 23*, 421–462. http://dx.doi.org/10.1093/applin/23.4.421

Dudeney, G., & Hockly, N. (2012). ICT in ELT: How did we get here and where are we going? *ELT Journal, 66*(4), 533-542. http://dx.doi.org/10.1093/elt/ccs050

Erten, I. H. (2014). The relationship between academic self-concept, attributions, and L2 achievement. *System*, *42*, 391–401. http://dx.doi.org/10.1016/j.system.2014.01.006

Field, A. (2009). *Discovering statistics for SPSS* (3rd ed.). Los Angeles, CA: SAGE Publications.

Ghonsooly, B., Khajavy, G. H., & Asadpour, S. F. (2012). Willingness to communicate in English among Iranian non–English major university students. *Journal of Language and Social Psychology, 3,* 197-211. http://doi.org/10.1177/0261927X12438538.

Gibson, D., & de Freitas, S. (2015). Exploratory analysis in learning analytics. *Technology, Knowledge and Learning*, *21*(1), 5–19.

Goggins, S. P., Galyen, K. D., Petakovic, E., & Laffey, J. M. (2016). Connecting performance to social structure and pedagogy as a pathway to scaling learning analytics in MOOCs: an exploratory study. *Journal of Computer Assisted Learning*, *32*(3), 244-266.

Hockly, N. (2013). Interactive whiteboards. *ELT Journal, 67*(3), 354-358. http://doi.org/10.1093/elt/cct021

Long, P., & Siemens, G. (2011). Penetrating the fog, analytics in learning and education. *Educause Review, 46*(5), 31-40.

Macfadyen, L., P., & Dawson, S (2010). Mining LMS data to develop and "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588-599. http://dx.doi.org/10.1016/j.compedu.2009.09.008.

Mah, D. K. (2016). Learning analytics and digital badges: potential impact on student retention in higher education. *Technology, Knowledge and Learning*, *21*(3), 285-305.

McCroskey, J. C., & Richmond, V. P. (1987). Willingness to communicate and interpersonal communication. In J. C. McCroskey & J. A. Daly (Eds.), *Personality and interpersonal communication* (129-156). Beverly Hills, CA: Sage.

Peng, J. E. (2011). Towards an ecological understanding of willingness to communicate in EFL classrooms in China. *System, 40*, 203–213. http://dx.doi.org/10.1016/j.system.2012.02.002.

Rienties, B., Hernandez Nanclares, N., Hommes, J., & Veermans, K. (2014). Understanding emerging knowledge spillovers in small-group learning settings: a networked learning perspective. In V. Hodgson, M. De Laat, D. McConnell, & T. Ryberg (Eds.). *The Design, Experience and Practice of Networked Learning* (Vol. 7, pp. 127-148). Dordrecht: Springer.

Sclater, N., A. Peasgood, & J. Mullan (2016). Learning analytics in higher education: A review of UK and international practice. Technical Report April, Jisc, London.

Shum, B., S., & Crick, D., R. (2012). Learning Dispositions and Transferable Competencies: Pedagogy, Modelling and Learning Analytics. *Proc. 2nd International Conference on Learning Analytics & Knowledge*, (29 Apr-2 May, Vancouver, BC). ACM Press: New York

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157-167.

Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. *Educational Technology & Society, 15*(3), 133-148.