FERİDE NUR HASKARACA BELIEF-RELATED WORDS AND THEORY OF MIND Bilkent University 2019

TURKISH SPEAKERS' CONCEPTUALIZATION OF BELIEF-RELATED WORDS AND ITS IMPLICATIONS FOR THEORY OF MIND DEVELOPMENT

A Master's Thesis

by

FERİDE NUR HASKARACA

Department of Psychology

İhsan Doğramacı Bilkent University

Ankara

May 2019

To My Mother, Handan Haskaraca

and

My Father, Mehmet Fatih Haskaraca

TURKISH SPEAKERS' CONCEPTUALIZATION OF BELIEF-RELATED WORDS AND ITS IMPLICATIONS FOR THEORY OF MIND DEVELOPMENT

The Graduate School of Economics and Social Sciences
of
İhsan Doğramacı Bilkent University

by

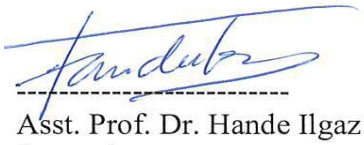
FERİDE NUR HASKARACA

In Partial Fulfillment of the Requirements for the Degree of
MASTER OF ARTS IN PSYCHOLOGY

THE DEPARTMENT OF
PSYCHOLOGY
İHSAN DOĞRAMACI BİLKENT UNIVERSITY
ANKARA

May 2019

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Psychology.



Asst. Prof. Dr. Hande Ilgaz
Supervisor

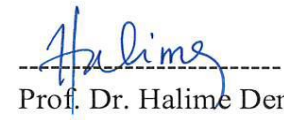
I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Psychology.

Asst. Prof. Dr. Jedediah Wilfred Papas Allen
Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Arts in Psychology.

Prof. Dr. Sibel Kazak-Berument
Examining Committee Member

Approval of The Institute of Economics and Social Sciences



Prof. Dr. Halime Demirkan
Director

ABSTRACT

TURKISH SPEAKERS' CONCEPTUALIZATION OF BELIEF-RELATED WORDS AND ITS IMPLICATIONS FOR THEORY OF MIND DEVELOPMENT

Haskaraca, Feride Nur
MA. Department of Psychology
Supervisor: Asst. Prof. Dr. Hande Ilgaz

May 2019

This thesis is comprised of three studies. Study 1 & 2 investigate whether there are pragmatic nuances between belief-related mental state verbs (e.g., "*to think, guess, and falsely think*") acknowledged by Turkish-speaking adults, and whether Turkish adults' implicit processing of the belief-including situations such as false belief tasks are affected by the appropriate (vs. inappropriate) use of mental state verbs. Study 3 investigates whether Turkish-speaking preschooler's performance in belief-related tasks of Theory of Mind (ToM) Battery [Diverse Belief (DB) and False Belief (FB) tasks, devised by Wellman & Liu, 2004] is affected by the verb used in these tasks. In Study 1, 150 Turkish-speaking adults completed an online survey asking for their judgments of appropriateness regarding the use of mental state words in belief tasks. In Study 2, 61 Turkish-speaking adults' accuracy rates and reaction times in response to interchangeable use of mental state verbs (MSVs) in FB

tasks were investigated. In Study 3, 60 Turkish-speaking children were tested on both the original ToM Battery and on the pragmatically modified versions of the DB and FB tasks. The DB and FB tasks were modified by either a) replacing the MSV used in the task (i.e., "*think"*) with a pragmatically and semantically more appropriate one (e.g., "*guess"* or "*falsely think"*); or, b) changing the epistemological circumstances of the task by adding an evidential basis for the belief so that the MSV used in the task (i.e., "*think"*) would be in line with the pragmatics of Turkish. Results revealed that Turkish-speaking children benefited from one modification that did not involve a manipulation of MSVs but the epistemological circumstances of the MSV (i.e., DB task presented with evidence).

Key words: Belief-Related Theory of Mind Tasks, Belief Words and Verbs, Pragmatics of Turkish Language, Theory of Mind

ÖZET

TÜRK KONUŞMACILARIN KANI İLE ALAKALI KELİMELERİ KAVRAMSALLAŞTIRMASI VE BU KAVRAMSALLAŞTIRMANIN ZİHİN KURAMI GELİŞİMİ İÇİN ÇIKARIMLARI

Haskaraca, Feride Nur
Yüksek Lisans, Psikoloji Bölümü
Tez Danışmanı: Dr. Öğr. Üyesi Hande Ilgaz

Mayıs 2019

Bu tez 3 çalışmadan oluşmaktadır. Çalışma 1 & 2, kanı ile alakalı kelimeler (örneğin "*düşünmek*", "*tahmin etmek*", "*zannetmek/sanmak*") arasında Türkçe konuşan yetişkinler tarafından kabul edilen edimsel (pragmatik) nüanslar olup olmadığını ve Türkçe konuşan yetişkinlerin kanı içeren durumları (örneğin yanlış-kanı görevleri gibi) işlemesinin bu kelimelerin doğru veya yanlış kullanımından etkilenip etkilenmediğin araştırmaktadır. Çalışma 3 Türkçe konuşan okul-öncesi yaştaki çocukların kanı içeren Zihin Kuramı (ZK) görevlerindeki [Farklı Kanılar (FK) ve Yanlış Kanı (YK), Wellman ve Liu, 2004] performanslarının bu görevlerde kullanılan kelimelerden etkilenip etkilenmediğini araştırmaktadır. Çalışma 1'de Türkçe konuşan 150 yetişkin, zihin durumunu ifade eden kelimelerin kanı görevlerindeki kullanımının uygunluğunu değerlendirmelerini isteyen çevrimiçi bir anket tamamlamışlardır. Çalışma 2'de Türkçe konuşan 61 yetişkinin, zihin durumunu ifade eden kelimelerin

birbirinin yerine kullanıldığı YK görevlerindeki doğruluk dereceleri ve tepki süreleri araştırılmıştır. Çalışma 3'te, Türkçe konuşan 60 çocuk, orijinal ZK görevleri ve FK ve YK görevlerinin modifiye edilmiş versiyonlarıyla test edilmiştir. FK ve YK görevleri iki şekilde modifiye edilmiştir: a) görevde kullanılan zihin durum ifadesi ("*düşünmek*") edimsel olarak daha uygun olan bir zihin durum ifadesi (örneğin "*tahmin etmek*" veya "*zannetmek/sanmak*") ile değiştirilmiştir; veya b) görevin epistemolojik durumu delile dayanan temellerle desteklenmiş, böylelikle "*düşünmek*" kelimesinin kullanımı Türkçe'nin edimsel özellikleriyle uygun hale gelmiştir. Bulgular Türkçe konuşan çocukların zihin durum ifadelerinin manipüle edildiği modifikasyonlardan değil, epistemolojik koşulların modifiye edildiği bir görevden (kanıt ile sunulan FK görevi) fayda sağladığını göstermiştir.

Anahtar kelimeler: İnanç İfade Eden Kelimeler ve Fiiller, İnanç ile Alakalı Zihin Kuramı Görevleri, Türk Dilinin Edimsel Özellikleri, Zihin Kuramı

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to my dear supervisor, Dr. Hande Ilgaz. This work would not have been possible without her patience and guidance. Without her knowledge and support, I would have never reached my current level. She has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general. As my teacher and supervisor, she has taught me more than I could ever give her credit for here.

I am also especially grateful to Dr. Jedediah Allen for his invaluable contributions to this study and for his lectures and feedbacks that increased my intellectual ability in general. I am also thankful to Dr. Sibel Kazak Berument for her valuable suggestions and comments on this study. I am grateful to Dr. Miri Besken for allowing me to use the facilities of cognitive psychology labs in Bilkent University.

I would especially like to thank İmge Saltık for her wonderful drawings that were used as testing materials and Elif Cemre Solmaz for her patience and help that made the second study of this thesis possible. I would also like to thank İrem Öztürk for her contribution to data entry.

I am grateful to the all members of Bil-Ge Lab in Bilkent University for their help and support. I could not have imagined working in a better lab. I would like to thank Bahar Bozbıyık and Elif Bürümlü Kısa for sharing their experiences with me whenever I needed and Eda Önoğlu Yıldırım and Aslı Yasemin Bahar for their endless support. I am also greatly thankful to the all participants for their invaluable contributions to this study.

I am very thankful to my dear friends, Esra Şenol and Elif Dirik whose love and support are with me in whatever I pursue.

I would like to express my love and gratitude to my parents, my brothers, and my sisters for their unfailing emotional and financial support. Without their support and trust, I would have never completed this work. I am also grateful to my nephew and my nieces who provide unending inspiration.

The last but not the least, I wish to thank my loving and supportive fiance, Hamza Kızılay. Without his endless love and support, it would be very difficult to overcome this process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Theory of mind (ToM) is the term used for children's ability to attribute mental states to themselves and to other people and to explain and predict behavior. This socio-cognitive ability shows major advances during the preschool years. Research on ToM has occupied socio-cognitive development literature for the past thirty years (e.g., see for reviews, Flavell, 2007; Flavell & Miller, 1998; Miller, 2012; Perner, 1991; Wellman, 1992; Wellman, 2014). Although understanding people as mental agents requires a cluster of interrelated abilities, early research on the development of ToM focused almost exclusively on understanding the false beliefs of others. Consequently, children's ToM abilities were predominantly measured by false belief paradigms (see for a review, Wellman, Cross, & Watson, 2001). False belief (FB) understanding is indeed a significant aspect of ToM development (see for theoretical and empirical works, Dennett, 1979; Flavell & Miller, 1998), yet, it is not possible to judge the developmental level of children's general ToM abilities just by looking at their success on FB tasks. ToM encompasses the attribution of many other mental states, such as desires, true and diverse beliefs, emotions, knowledge levels, intentions, etc. (Wellman, 2002). The multi-component structure of ToM created a need for a comprehensive set of tasks to measure the development of ToM abilities. In 2004, Wellman and Liu came up with a comprehensive ToM battery that satisfied

the need to a large extent. In addition to assembling tasks that measure different components of ToM, Wellman and Liu (2004) also proposed that the understandings measured by these tasks develop in a predictable order which would be seen by scaling ToM tasks, (i.e., sequencing the tasks according to performance).

Many other scaling studies followed the seminal work of Wellman and Liu and the translations of the same ToM tasks were applied on preschoolers from diverse countries such as Germany, Australia, China, Iran, Indonesia, and Turkey (Etel & Yağmurlu, 2015; Selçuk, Brink, Ekerim, & Wellman, 2018; Kristen, Thoermer, Hofer, Aschersleben, & Sodian, 2006; Kuntoro, Saraswati, Peterson, & Slaughter, 2013; Shahaeian, Peterson, Slaughter, & Wellman, 2011; Wellman, Fang, Liu, Zhu, & Liu, 2006; Wellman & Liu, 2004). These studies, in general, found that although children's overall performances on the scale did not show much variation between cultures, there was variation in the developmental progression of ToM tasks by culture. Specifically, it was found that children from Western societies such as the Unites States (US), Germany, and Australia acquired diverse belief understanding before appreciating that perception leads to knowing (i.e., knowledge access). This was termed the Western Pattern. In contrast, children from Eastern societies such as China and Iran, passed knowledge access tasks before diverse belief tasks. This has been dubbed as the Eastern pattern. The two scaling studies (i.e., Etel & Yağmurlu, 2015; Selçuk, et al., 2018) conducted with Turkish children yielded conflicting results regarding the developmental pattern of ToM acquisition: the earlier study suggested the Western pattern while the later one suggested the Eastern pattern.

However, we should note that the earlier study (Etel & Yağmurlu, 2015) was conducted with institutionally-reared Turkish preschoolers who were likely to show different developmental progressions than their family-reared counterparts. The later study (Selçuk, et al., 2018) includes a typical sample and hence is more telling of the general pattern followed by Turkish children. In this study, Selçuk and colleagues found that Turkish children followed the same pattern as the Chinese and Iranian samples. A third study was conducted by the Bil-Ge Lab in Bilkent University (Ilgaz, Allen, & Haskaraca, 2019) and preliminary results support the finding that Turkish preschoolers fit in the Eastern pattern. Besides the pattern found for Turkish children, what was actually striking about these studies is that they demonstrated that Turkish children were scoring dramatically lower in belief-related task (diverse and false beliefs) of Wellman and Liu's (2004) ToM battery compared to their peers from the US, Australia, and China. We argue that using the direct translations of the tasks (e.g., Kahraman, 2012; Özoran, 2009) that were originally created for English-speaking children may ignore the pragmatics associated with mental state verbs in Turkish. Specifically, we argue that, small but significant nuances in the way Turkish speakers choose to use the verb "*think*" may be the reason for Turkish children's lower performance in ToM tasks using this verb.

In the current study, we aimed to investigate how belief-related tasks in the ToM battery could be modified so that they become more appropriate for Turkish's linguistic characteristics. Depending on the literature and our knowledge of Turkish, we modified the belief-related tasks: we either replaced the MSV used in the task (i.e., "*think"*) with a pragmatically and semantically more appropriate one (e.g., "*guess"* or "*falsely think"*) or we changed the way task was presented so that the MSV used in

the task (i.e., "*think"*) would be in line with the pragmatics of Turkish. Once we created these modifications, we wanted to ensure that they were successfully representing the more accurate uses of MSVs in Turkish. For this purpose, we conducted 2 studies with Turkish-speaking adults. In Study 1, we aimed to get insight about Turkish adults' explicit judgments of appropriateness of different MSVs. In Study 2, we investigated whether the use of different MSVs affected Turkish adults' performance in FB tasks. In Study 3, we administered the modified versions of belief-related tasks along with their original versions to Turkish children to examine whether Turkish children benefited from the modifications. Finally, within the scope of Study 3, we sought to investigate the sequence of development with the original and the modified versions of the tasks.

In the following sections, Wellman and Liu's (2004) ToM battery and the other studies using the same battery, or its direct translation, will be examined. While examining these studies, we will compare ToM scores of children coming from diverse cultures in terms of both total score and tasks' sequences and we will make arguments about the different patterns. Then, we will point to the relatively low scores of Turkish preschoolers in belief-related tasks. We will present what we thought would be the reason for Turkish children's underperformance in belief-related tasks which are mostly about the tasks' lack of congruence with Turkish's pragmatics. Finally, we will propose possible modifications to overcome these incongruences and present our hypotheses regarding the effects of these modifications.

1.1 Wellman and Liu's (2004) ToM Scale and American Children's Performance on This Scale

Theory of mind (ToM) is a complex socio-cognitive ability entailing an understanding one's own and others' various mental states such as desires, emotions, beliefs, knowledge states (Wellman, 1992). Although this construct includes a variety of mental components, the early research on preschoolers' ToM development was mostly shaped around preschoolers' performance in FB paradigms since false belief understanding was accepted as the major, or sometimes definitive, indicator of mental state understanding for both typically (e.g., Astington & Jenkins, 1999; Dunn, Brown, Slomkowski, Tesla, & Youngblade, 1991; Hughes & Dunn, 1998; Lalonde & Chandler, 1995; Perner, Ruffman, & Leekam, 1994; Youngblade & Dunn, 1995) and atypically developing preschoolers (e.g., Baron-Cohen, 1995; Baron-Cohen, Leslie, & Frith, 1985). The reason for the importance given to false belief understanding was nicely summarized by Wellman and his colleagues. Based on theoretical works about false beliefs (e.g., Dennett, 1979), authors claimed that:

> Mental state understanding requires realizing that such states may reflect reality and may be manifest in overt behavior, but are nonetheless internal and mental, and thus distinct from real-world events, situations, or behaviors. A child's understanding that a person has a false belief -one whose content contradicts reality- provides compelling evidence for appreciating this distinction between mind and world (Wellman, Cross, and Watson, 2001, p. 655).

However, while false belief understanding is a critical component of the ToM construct, it was also necessary to systematically document the developmental

trajectory of supporting ToM abilities. The literature showed developmental gains in related ToM abilities such as appreciating divergent desires, emotions based on desires, diverse beliefs, and the link between perception and knowledge (e.g., Fabricius & Khalil, 2003; Friedman, Griffin, Brownell, & Winner, 2001; Olson, Liu, Kerr, & Wellman, 2003; Ruffman, Slade, & Crowe, 2002). Importantly, it was also shown that children's ToM competence was not complete with false belief understanding. For instance, children's understanding of false beliefs seems to precede their understanding of emotion-based beliefs and understanding that expressed emotions can differ from the real emotions (Harris, Donnelly, Guz, & Pitt-Watson, 1986; Vinden, 1999). Therefore, depending solely on false-belief paradigms cannot tell the whole story of the development of mental state understanding. To that end, tasks measuring various ToM abilites should be used together in order to get the most accurate and comprehensive picture of ToM development. The first, and probably the most prominent, effort on this issue was made by Wellman and Liu in 2004. The seminal work of the authors provided us with a widely used ToM battery.

While creating the ToM battery, Wellman and Liu (2004) did not aim just to gather tasks that tap into different aspects of ToM; rather, they intended "to assemble a set of tasks that are easier or harder because of conceptual differences among them." (p. 530) and to scale these tasks so that they could reflect the acquisition order of the underlying concepts. The effort to create a set of tasks that can measure the changes in the process of acquiring mental state understanding is a reflection of Wellman's Theory Theory of ToM.

Theory Theory is one of the theories that have been proposed to explain the emergence of children's mental state understanding. The central claim of this theory

is that understanding minds emerges in a theory-like manner. Through social interactions and language, children construct, and revise theories related to mind and how it works. Initially, children are "desire psychologists", i.e., they develop theories about other's behavior by understanding their desires. Then, they become "belief psychologist" and explain the behavior on the basis of beliefs. Children achieve a progression from initial concepts to intermediate understandings as initial conceptions and theories fail to explain behavior and perceptually accessible events (Gopnik & Wellman, 1992; Wellman, 1990; Wellman; 2014).

In order to assemble the tasks and to gain insight about the acquisition order of the various abilities, at first, Wellman and Liu conducted a meta-analysis and summarized the earlier findings on the differences between various mental state understandings. Then, based on the results of the meta-analysis, they created a 7-item ToM battery, applied it to English-speaking American preschoolers and eventually settled on a 5-item ToM scale in which tasks were consistently sequenced (see Table 1 for task descriptions).

Table 1. Description of ToM Tasks

Tasks	Task Descriptions
Diverse Desires (DD)	*"Child judges that two persons have different desires about the same objects."*
Diverse Beliefs (DB)	*"Child judges that two persons have different beliefs about the same object, when the child does not know which belief is true."*

Table 1 (cont'd)	
Knowledge Access (KA)	*"Child sees what is in a box and judges the knowledge of another person who does not see what is in a box."*
Content False Belief (CFB)	*"Child judges another person's false belief about what is in a distinctive container when the child knows what is in the container."*
Explicit False Belief (EFB)	*"Child judges how someone will search, given that person's mistaken belief."*
Hidden Emotion (HE)	*"Child judges that a person can feel one thing but display a different emotion."*

Note. Task descriptions are quoted from Wellman & Liu, 2004, p. 531.

1.1.1 Meta-Analysis

This analysis was done in order to investigate the earlier studies which compared one kind of mental state understanding to another and to use their findings for the selection of tasks to be used in the primary study. The authors were interested in 4 type of comparisons: between diverse beliefs and false beliefs, between diverse desires and diverse beliefs, between satiated desires and outdated -hence false- beliefs, and between knowledge and false beliefs. Results revealed that children accurately judged other's desires earlier than their beliefs. Similarly, and even more consistently, children correctly judged diverse beliefs earlier than false beliefs. Contrary to the previous literature that found inconsistent results, the meta-analysis showed a consistent difference between understanding knowledge access and false beliefs: children were correctly estimating others' knowledge state before they could judge their false beliefs. These findings informed the authors' task selection for the

primary study and also supported the idea of scalable ToM tasks which are passed progressively more as children get older.

1.1.2 American children's performance on the ToM scale

The aim was to create a ToM scale that can address the progression in mental state understanding by investigating children's performance on individual tasks. Wellman and Liu believed that, albeit being similar in terms reflecting subjectivity, certain understanding related to mind are conceptually different than each other and they develop in a predictable order and form a developmental progression that can be measured by the ToM scale. Based on the preliminary findings of the meta-analysis, the primary study included 5 tasks assessing Diverse Desires, Diverse Beliefs, Knowledge Access, and False Beliefs (two tasks: Contents FB, Explicit FB). Even though emotion understanding was not compared to any type of mental state understanding due to task differences in the meta-analysis, the primary study also included 2 tasks assessing emotion understanding (Belief-Emotion and Hidden Emotion) in order to be comprehensive. It resulted in a 7-item ToM Scale in which each item (except two FB tasks) targeted a conceptually different understanding (see Table 1 for brief descriptions of the tasks). The authors also stated that they tried to keep the tasks similar in terms of non-relevant task-performance requirements. However, the tasks were still not comparable in some aspects such as number of control questions, the length of the task vignette, and the structure of the question. In order to overcome this issue, they a) analyzed the data by using a number of analyses, and b) they included two different FB tasks which were supposed to reveal comparable results and to scale similarly.

The 7-item ToM battery was applied on 75 English-speaking American children who were either 3-, 4-, or 5-years old. First, children's overall performance on ToM scale was analyzed in order to reveal the effects of age, gender, and tasks' presentation order. It only revealed a significant effect for age, i.e., as children got older, they passed more tasks. Then Guttman and Rasch analyses were conducted in order to scale the tasks according to children's performances on them. Guttman analyses (will be explained in Chapter 4) is a more stringent scaling analysis and it revealed that 5 tasks in the battery were highly scalable: Diverse Desires (DD), Diverse Beliefs (DB), Knowledge Access (KA), Content False Belief (CFB) and Hidden Emotions (HE), from easiest to hardest (see Table 2 for children's performance on each ToM task). The Explicit False Belief (EFB) task was not included among the scalable items since the performance on two FB tasks were essentially equivalent. This task (i.e., EFB) was included in some of the later scaling studies but not in Wellman and Liu's original scaling analysis. In addition, the Belief-Emotion task was not applied in studies that followed the original Wellman and Liu (2004) study and it was never put in the scaling analysis since the original study showed that children's performances on this task did not yield scalable results (therefore, it is not included in Table 1). Finally, task comparisons supported the order revealed by Guttman analysis. Accordingly, DD was easier than DB, DB was easier than KA, KA was easier than CFB, and CFB was easier than HE.

Wellman and Liu concluded that the results of Study 2 validated the ToM scale they created. They argued that this scale is especially important for two reasons: a) it demonstrated the progression of conceptual achievements indicating the socio-cognitive abilities of typically-developing preschoolers, and b) it is a comprehensive

battery that measures ToM development accurately. These two arguments will be at the center of the current study. We will take a socio-cognitive perspective and investigate whether the conceptual progression is identical and discuss whether linguistic and cultural factors may impact children's understanding of the subcomponents of ToM. To that end, we will investigate whether both arguments hold in a culturally different sample as compared to the English-speaking samples using original and modified versions of the tasks.

1.2 Cross-Cultural Differences in Children's ToM Performances

Wellman and Liu's pioneering work on scaling ToM tasks attracted immediate attention and many other studies using the same ToM scale were conducted with preschoolers from diverse cultures and sub-cultures. These studies varied highly in their purposes for using the ToM scale. Some of them used the scale just to measure ToM understanding of children comprehensively and to study their overall ToM ability (e.g., Devine & Hughes, 2016; Meins, Fernyhough, Arnott, Leekam, & Rosnay, 2013) whereas others aimed to scale the ToM tasks for preschoolers residing in different cultures and speaking different languages (e.g., Shahaeian et al., 2011; Wellman et al., 2006). The first group of studies were interested in total score and did not report children's performances on individual tasks; therefore, they did not provide the most useful information for our cross-cultural comparisons of single tasks. In contrast, scaling studies conducted with children from diverse cultures were more informative about the differences between children's performances in individual tasks along with their total ToM scores and the sequence of ToM tasks. Therefore, we will exclusively focus on scaling studies in the following sections, summarize and compare their results with the original scaling study. In cases of difference in results

for either the total ToM score children attain or the sequence of passing these tasks, we will briefly explain the authors' speculations about these differences. In addition, we will report if task modifications were included. These will be especially useful in situating the modification we made in the current study to the literature. At this point, we should note that nearly all scaling studies used the same analyses in order to scale ToM tasks, e.g., Guttman scale, Rasch analyses, and McNemar's chi-square tests. Therefore, we will not go over these analyses again unless there is a novel analysis; instead, we will report the general results of these studies.

1.2.1 Australian children's performance on ToM scale (Peterson, Wellman, and Liu, 2005)

One study aimed to investigate whether Australian children with autism or deafness followed a different developmental pattern when acquiring ToM tasks as compared to typically developing children. For comparison purposes, the authors tested typically developing Australian children who were between 3 ½ and 5 ½. Atypically developing children's performance is not related to this thesis, therefore, will not be discussed further. Typically developing, English-speaking Australian children who were tested on the 5 tasks of the original ToM battery not only performed similarly in total ToM ability, but also followed the same sequence that was observed earlier in the original study. Accordingly, the tasks consistently and significantly were scaled as follows: DD was easier than DB, DB was easier than KA, KA was easier than CFB, and CFB was easier than HE (see Table 2 for children's performance on each ToM task).

The tasks used in their study closely modeled the ones used in the original study with a few small changes: the wording of the tasks and the stimuli used in the tasks were

modified in order to a) simplify the task language so that it can be directly translated to sign language, e.g. shorter sentences used, and b) to make tasks more familiar for Australian children, e.g., in DD task, a biscuit was used instead of a cookie. A notable change was done in HE task: one of the control questions was excluded since it was too long and complicated to be translated into the sign language.

1.2.2 Chinese children's performance on ToM scale (Wellman, Fang, Liu, Zhu, and Liu, 2006)

The aim of another study was to investigate the extent to which the sequence of ToM understanding was universal. For this purpose, the authors administered Wellman and Liu's (2004) ToM scale on 3-, 4-, and 5-years old Mandarin-speaking Chinese preschoolers and compared their performances with their English-speaking American (data retrieved from Wellman and Liu, 2004) and Australian peers (data retrieved from Peterson et al., 2005) in terms of both total ToM scores and the sequence of ToM understanding. Results revealed that the three groups of children performed similarly in terms of total ToM score; however, the scale order of understanding DB and KA was switched for Chinese children. As distinct from American and Australian preschoolers, Chinese children acquired knowledge access understanding before acquiring diverse beliefs. The rest of the tasks were equally scalable in both cultures. Namely, DD was acquired in the earliest stage and it was always easier than understanding diverse beliefs or knowledge access. For both cultures, KA was easier than FB and HE was the hardest task. (see Table 2 for children's performance on each ToM task)

The difference between the patterns (i.e., DB and KA switching) was hypothesized to emphasize the sociocultural influences on ToM development. According to the authors, although children from different cultures work on universal set of ToM insights or understandings, these children receive different information pertaining to mental states and they experience mental states differently. These differences may result in a variance in the sequence of ToM tasks. For example, collectivistic, Confucian-oriented Chinese culture emphasizes knowing and practical knowledge for children whereas more Western societies like the US and Australia put more emphasis on diversity and truth of beliefs. Therefore, while knowing and knowledge become prominent in China, understanding diversity of people's belief is more salient in Western cultures. These differences between cultures' emphases can result in different sequences of ToM understandings as Chinese children acquire knowledge before beliefs and children from Western societies acquire beliefs before knowledge states of others.

Wellman and his colleagues' study used a parallel Mandarin version of the 6 tasks originally used in Wellman and Liu's (2004) study. Task protocols were translated and back-translated by English-Mandarin and Mandarin-English bilinguals in order to get culturally and linguistically appropriate tasks that were also conceptually equivalent to their original counterparts. As a result, the Mandarin tasks closely resembled the original tasks with small changes in materials and wording. Materials were switched to ones that were more familiar for Chinese children, e.g., the toy figures of the protagonists had dark hair and Chinese countenances; ice cream and egg were the desired objects in DD task, rather than cookie and broccoli; a potato-chip box, instead of a Band-Aid box, was used in CFB task; and HE task was about a boy

and his uncle, rather than a boy and his peers. When it comes to the wording of questions, there was a small but notable change: the target question of CFB task was asked with Chinese-specific "*think-falsely"* verb (i.e., *yi3wei2*). Earlier studies demonstrated that using this verb in FB tasks sounded more appropriate for Chinese children and therefore enhanced their performance in those tasks (e.g., Lee, Olson, & Terrance, 1999). Wellman and his colleagues (Wellman et al. 2006) did not want to withhold any opportunity from Chinese children to show their understanding of false beliefs and, therefore, switched to use the "*think-falsely"* verb in the target question of CFB task. The authors proposed no change for the wording of the EFB task given that the target question of the EFB task did not contain a MSV in its original; however, by doing this, they ignored the fact that the vignette of the EFB task included the "*think"* verb to refer to a situation in which "*think-falsely"* verb could have been more appropriate.

1.2.3 Iranian children's performance on ToM scale (Shahaeian, Peterson, Slaughter, and Wellman, 2011)

A fourth study aimed to investigate the cultural contrasts that were leading to sequential differences observed in the acquisition of ToM tasks. For this purpose, the authors compared 3-to-6-years old, middle-class Australian and Iranian children who speak English and Farsi, respectively, on the 5-item ToM scale developed by Wellman and Liu (2004). Results revealed similarity in terms of overall mastery in ToM; however, there was a cross-cultural difference in the ToM scales. Supporting the earlier findings (Peterson et al., 2005), Australian children followed the sequence that was originally found with American preschoolers whereas Iranian children followed the pattern that was previously found for Chinese preschoolers in which

diverse belief and knowledge access understandings switched places. Otherwise, the sequence of the tasks was similar for both cultures: DD was easier than DB/KA, DB/KA was easier than CFB, and CFB was easier than HE (see Table 2 for Iranian children's performance on each ToM task, Australian children's performance is presented under Peterson et al., 2005). The authors discussed that these findings are consistent with the importance given to acquiring knowledge, filial respect, and dispute avoidance by Iran's collectivistic culture as opposed to the importance given to different opinions and divergence of beliefs by more individualistic, Western cultures such as United States and Australia. Specifically, the authors claimed that children who are raised in collectivistic cultures such as China and Iran could be socialized to value knowledge of themselves and others by the means of warm-but-authoritarian parental practices. This type of socialization and certain parental practices that are common in China and Iran favor children's early appreciation of knowledge, therefore earlier mastery in the KA task. At the same time, children growing up in more individualistic, Western societies are encouraged to express and assert their opinions by their independence-oriented parents. These approaches adopted in Western cultures were claimed to favor children's early mastery in the DB task.

When it comes to task modifications applied in this study, in order to ensure that task protocols were closely comparable to each other to earlier studies, Australian children were tested on the original (English) version of the ToM scale whereas the scale was translated from English to Farsi and then back-translated (for control purposes) to test Iranian children. Three small changes were done in both versions of ToM scale, i.e., Farsi and English versions. First, one desired object in DD task and the descript box

in the CFB task were replaced with locally more familiar ones; a cake rather than a cookie and a candy box instead of a Band-Aid box were used. Second, children's memory burdens were reduced by naming the protagonists, rather than calling them *the girl* or *the boy*. And third, an additional *why* question was added to the HE task in order to probe children's reasoning in their answers to test questions.

There have been other studies measuring children's performances on ToM scale in various countries such as Germany and Indonesia (e.g., Kristen, et al., 2006; Kuntoro, et al., 2013), but we will not go into the details of these studies for several reasons. First of all, the study conducted with German children (i.e., Kristen et al., 2006) was published in German. Secondly, the study conducted with Indonesian children (i.e., Kuntoro et al., 2013) tested samples that were importantly different: typically-developing middle-class children and trash pickers. Although these children showed significant differences in their ToM performances, specifically in the KA and the HE tasks, the authors combined two groups when scaling the tasks. We believe this statistical practice may not produce comparable results and might result in reducing the validity and reliability of ToM scale. Thirdly, both studies used the exact translation of the original ToM battery except for minor changes made on the materials, therefore, they did not inform our selection of task modifications any further. Finally, none of these studies revealed an unusual pattern for the developmental progression of ToM tasks, i.e., both revealed Western pattern.

1.2.4 Turkish children's performance on ToM Scale (Etel & Yağmurlu, 2015; Selçuk, Brink, Ekerim, & Wellman, 2018)

To our knowledge, so far, two studies have investigated the sequence of ToM tasks for Turkish children and they yielded conflicting results. The general aim of the earlier study was to investigate ToM, executive functioning and social competence in institution-reared Turkish children. To this end, the sample consisted exclusively of institution-reared children. The direct translation of a 6-item ToM scale (see Kahraman, 2012) was applied on 3-to-5-years old children who were raised in home-like institutions where approximately 10 children were living together. Guttman analyses favored the Western pattern for institution-reared Turkish preschoolers in which the DB task was acquired before KA however, only 4 tasks including DD, DB, KA, and FB tasks were scalable (see Table 2 for children's performance on each ToM task). Specifically, HE did not differ from the FB tasks in terms of children's performance and it did not form a separate step in the scale. We should also note that since the performance on EFB and CFB tasks were not significantly different than each other, the authors used only the CFB task as the indicator of FB understanding. Although the authors found a consistently scalable pattern indicating the developmental sequence of mental state understanding, they also acknowledged that the pattern they found may not be generalizable to typical Turkish children from intact-family backgrounds. Given that children lived with large numbers of peers (as compared to family contexts where number of siblings would be more limited), institution-reared preschoolers could be exposed to diverse thoughts and opinions in their daily interaction with peers more frequently. This may result in increased competence in understanding of diversity.

Selçuk and colleagues' recently ran another study with typically developing Turkish children from intact family backgrounds (Selçuk, Brink, Ekerim, & Wellman, 2018). Three-to-6 ½ -years old children from 5 cities in Turkey were tested on the direct translations of the 5-item ToM scale (Wellman & Liu, 2004). Results revealed that Turkish children showed the Eastern pattern that was previously shown by Chinese and Iranian preschoolers: they acquired knowledge access earlier than diverse beliefs. Contrary to the earlier study's failure in scaling the HE task, all 5 tasks were found to be scalable in the later study (see Table 2 for children's performance on each ToM task).

In addition to usual scaling analyses, Selçuk et al. (2018) further analyzed their data to investigate if a subsample reflected a different sequential pattern. Given that Turkish culture is a unique mix of collectivistic and individualistic cultures, Selçuk and colleagues hypothesized that there could be a subgroup of children who follow the Western pattern when acquiring ToM tasks. In order to do this, they excluded the children who fit in the majority pattern (Eastern) from their data and ran the scaling analyses with the remaining 21 children. This re-analysis revealed that the remaining children fit the Western pattern in which DB is acquired earlier than KA. The authors were also interested in what distinguished the children who followed the two different patterns. In order to find the predictors of the two different patterns, they conducted logistics regressions and looked at whether the variables differing between two groups (age and number of adults in household) significantly predicted either pattern, i.e., DB before KA vs. KA before DB. Logistic regression analyses showed that DB before KA pattern (Western pattern) was predicted negatively by age whereas the number of adults at home positively predicted this pattern. The authors' discussion of these

results mostly focused on the individualism-collectivism dichotomy. For instance, they argued that the larger households where multiple ideas and viewpoints are likely to occur more frequently resemble the individualistic cultures in which different opinions are expressed freely and frequently. Since children living in larger households are more likely to be exposed to different ideas more frequently, just like their peers growing up in individualistic cultures, they acquire diverse belief understanding earlier than understanding knowledge (vs. ignorance). However, these results are not able to explain a) why just the number of adults, but not the number of siblings, in the household is contributing to the diverse belief understanding, and b) why DB is not acquired earlier in more traditional societies such as Iran where multiple adults (parents, grandparents, aunts and uncles) usually live together. We believe that, although it is not discussed further by the authors, these findings are especially important in terms of revealing the effect of age on children's performance on DB task and signaling that scaling analyses should consider the potential differences in the sequencing of ToM tasks caused by age.

A third study on scaling ToM tasks among Turkish children was conducted by our lab (Bil-Ge Lab: Bilkent Developmental Psychology Laboratory) (Ilgaz, Allen, & Haskaraca, 2019). This study used the direct translation (see Özoran, 2009) of Wellman and Liu's (2004) ToM scale with a large sample (N = 383). Preliminary results indicate that, if all age groups are combined, Turkish children seem to fit in the Eastern pattern of ToM tasks: they acquire knowledge access before diversity of beliefs (see Table 2 for children's performance on each ToM task). However, when age groups were examined more closely, variability was found between the patterns followed by each age group. The patterns were not completely different than each

other and they were somehow related: the KA task became consistently easier as age increased and its place on the scale switched from third to first from 3-to-5 years of age. Logistic regression analysis revealed that children's EF abilities predict their success in KA task above and beyond children's age. In other words, children who had higher EF abilities were more likely to succeed in the KA task. This finding suggested additional scaling analyses by separating children into groups according to their performance on the EF task: low, middle, and high. This helped us investigate the effect of EF abilities on the scaling of ToM tasks. As a result, we found that children with low EF abilities fit in the Western pattern in which DB task is acquired later than the DD but earlier than the KA task. Children with average EF abilities, however, fit in the Eastern pattern and they acquired the KA task before DB. Finally, children with high EF abilities, fit in the pattern that we previously found for 5-years old Turkish children: they acquired the KA task first and it is followed by DD and DB tasks. These results are adding to the findings of Selçuk et al., (2018) who showed the effect of age on the scaling of DB and KA tasks and further suggests that age should be considered simultaneously with other cognitive variables.

Table 2. The Percentage of Children who Passed Each ToM Task by Study

Tasks	Wellman & Liu, 2004	Peterson et al., 2005	Wellman et al., 2006	Shahaeian et al., 2011	Etel & Yağmurlu, 2015	Selçuk et al., 2018	Bil-Ge Lab
	American Children	Australian Children	Chinese Children	Iranian Children	Turkish Children		
DD	95%	95%	89%	86%	91%	82%	82%
DB	84%	85%	71%	47%	71%	64%	57%
KA	73%	82%	79%	88%	44%	79%	73%

Table 2 (cont'd)							
CFB	59%	32%	54%	16%	12%	24%	38%
EFB	57%	-	49%	-	20%	-	21%
HE	32%	19%	37%	17%	19%	19%	19%

Note. Reported results are from typically developing preschoolers.

1.2.5 Summarizing the Cross-Cultural Differences in ToM Scale

The 6 scaling studies reviewed in detail indicate sequence differences across cultures. These studies have revealed two different ToM patterns which are believed to vary based on culture's classification according to the individualism vs. collectivism dichotomy: a) DB is mastered earlier than KA in individualistic Western cultures where diversity of beliefs and individual opinions are valued; b) KA is mastered earlier than DB in more traditional, collectivistic Eastern cultures where possessing knowledge (vs. ignorance) and accessing shared knowledge, rather than having individual opinions, are valued (Wellman et al., 2006; Shahaeian et al., 2011, Selçuk et al., 2018). Further, previous studies have suggested sequence differences based on sub-culture, such as home environment (Etel & Yağmurlu, 2015 vs. Selçuk et al., 2018), SES differences (trash pickers vs. middle-class children, Kuntoro et al., 2013), and deaf children with hearing or deaf parents (Peterson et al., 2005). Despite the variability, the reasons for sequence differences have not received much attention. Most of the extant research with Wellman & Liu (2004) scale have aimed to highlight the universality of ToM development, and where cultural differences are evident their explanations have remained as speculations rather than viable research questions.

In addition to scale differences, what is intriguing about the data presented in Table 2 is that although the prominent acquisition sequence is similar to other cultures such as China and Iran, Turkish children tend to score lower in the DB task then their peers from China. It is true for Turkish children who show earlier mastery in the DB task than the KA task (Etel & Yağmurlu). Even then, Turkish children score lower in the DB task than their peers from the US and Australia, who follow the same ToM sequence. A similar underperformance for Turkish children is observed in the FB tasks. Despite the fact that its place on the sequence never changes, Turkish children score lower in FB tasks (especially in the EFB task) compared to their peers from the US, China, and Australia. Interestingly, Iranian children also show comparably low DB and FB performances. It is possible that socio-cultural factors that shape the way we think and talk about mental states may be at play in cross-cultural performance differences in ToM. The current study aimed to examine the pragmatic factors that may affect Turkish children's ToM performance. Given that our research team is unfamiliar with Iranian culture in general and Farsi in particular, we will not be able to extend our results to this sample.

In our opinion, the individualism vs. collectivism dichotomy may not be able to explain all of these results. If Turkey is accepted as dominated by collectivistic features, we would expect Turkish children's performance to be similar to their peers from China, the representative culture for collectivism in the ToM literature. However, as stated previously, Turkish children (and Iranian children) lag behind their Chinese peers in both the DB and the FB tasks. Another alternative explanation is that Turkish children are especially bad at understanding beliefs. Such an alternative may be unlikely given the diversity of cultural richness in the Turkish

population and the need for nuanced understanding of others' beliefs in order for the community to function (Tamer-Gencer, 2011).

We believe that there is an alternative explanation that may account for Turkish children's underperformance in belief-related tasks (DB and FB tasks) of Wellman and Liu's (2004) ToM scale. This explanation is based on socio-cultural view where cultures provide individuals with ways to think and talk about mental states in relation to behavior. We will argue that, nuances in the way Turkish-speakers express belief-related states vary in minor yet impactful way as compared to English speakers' general intuitions about mental states. In the following section, we will present our argument and its implications for how it may affect Turkish children's performances in belief-related tasks.

1.3 The Effect of Task Language on Turkish Children's Performance on Belief-Related ToM Tasks: Are we lost in translation?

An alternative view to Wellman's Theory Theory asserts that children are socialized into their communities' ways of construing the mind (e.g., Nelson, 1996). From this perspective while there are universals of social cognition since human beings in groups act in predictable ways, there is inevitably nuances based on how communities think and talk about the mind. Importantly, these nuances would have developmental consequences.

In the current study, we aimed to explore whether such differences could explain young children's performances on belief-based tasks. Specifically, we suspected that the direct translations of tasks may have produced task instructions that are not congruous with how adults talk about beliefs and false beliefs to children (and to one

another) in the Turkish culture. We investigated whether the use of the belief verbs "*think*", "*guess*", and "*falsely think*" have a differential impact on children's performances on belief-related ToM tasks.

In the original, English version of Wellman and Liu's ToM scale, belief-related tasks (DB and FB tasks) include the "*think"* verb either in the vignette of the task or in the target question and protagonist's divergent/false beliefs is presented with this verb. In order to preserve similarity to the original tasks, later scaling studies including the ones conducted with Turkish children, adhered to the direct translation of these tasks. Therefore, they used the direct translation of "*think"* in the relevant language (e.g., düşünmek, in Turkish). In our opinion, the verb "*think"* in Turkish (i.e., düşünmek) is distinct in both its pragmatics and semantics from some of the other relevant belief verbs (i.e., "*guess"* and "*falsely think"*) with small but meaningful nuances.

First of all, Turkish emphasizes evidentiality through explicit markers. That is, Turkish is one of the few languages that obliges its speakers to explicitly mark whether they have first-hand evidence for a particular claim (Aksu-Koç, 1988; 1998; Aksu-Koç, Ögel-Balaban, & Alp, 2009). The "–mIş" suffix is thus used to indicate either a claim that the speaker has not experienced firsthand (e.g., *Dün Ankara'da yağmur yağ-mış*. "I was told that it rained in Ankara yesterday"), or one in which the statement is based on an evidence-based inference rather than firsthand knowledge (e.g., *Annem eve gel-miş*. "Through seeing mum's shoes in front of the door, I infer that she is at home). Relatedly, the verb "*think*" seems to be used in a more factive and evidence-based manner or in instances where there is reason to hold a particular belief. In all other instances, the verb "*guess*" may be more appropriate.

Moreover, since Turkish puts greater emphasis on evidentiality and explicitly marks it in everyday talk, Turkish children might be more sensitive to a speakers' basis for knowledge. In their review of the developmental trajectory of evidentiality in Turkish-speaking children, Aksu-Koç et al. (2009) state that although young children start using the evidential suffix from the second year of life, they do not fully appreciate the different functions (e.g., inference vs. second-hand knowledge) in production tasks until 4 years of age and in comprehension tasks until 5 years of age. Further, it is only by 6 years of age that children show a nuanced understanding and can provide explicit reasoning for their choices between evidential vs. the direct-experience marker. Regardless of when Turkish children show full competence in these epistemologically distinct expressions, in everyday child-directed talk they hear these expressions abundantly. It is plausible to assume that Turkish children who are in the process of consolidating their understanding of different degrees of knowing would be sensitive to the use of different mental state verbs (MSVs) in reference to different knowing states. This sensitivity may not reveal itself in full-fledged competence but may manifest as expectations for children's interpretations of the task instructions that they hear.

In contrast, English-speaking children hear interchangeable usage of epistemologically nuanced belief-indicating verbs, especially "*think"* and "*guess"*. In turn, studies with English-speakers have shown that children had a difficult time differentiating these verbs by as late as 8-years of age (Moore, Bryant, & Furrow, 1989; Moore, Pure, & Furrow, 1990). Similarly, English-speaking children do not appreciate the distinction in the amount of certainty expressed by "*know"* vs. "*guess"* and by "*know"* vs. "*think"* verbs until 4 years of age (Macnamara, Baker, & Olson,

1976; Miscione, Marvin, O'Brien, & Greenberg, 1978; Johnson & Maratsos, 1977; Johnson & Wellman, 1980). Due to the fact that these MSVs are used interchangeably English-speaking children extends the developmental time frame of their full appreciation, allowing them to gradually come to differentiate the levels of uncertainty they express. However, this also implies that the standard versions of ToM tasks where the verb "think" is used without an evidential rationale may not adversely affect English speaking children's performance. To be specific, an English-speaking child may not find it odd when a question asks where a cat could be hiding and gives two options. Although the child does not have any idea about the location of the cat, any choice between location 1 and location 2 might be expressed using the verb "*think*" (e.g., *I think the cat is in the bushes*).

Our argument is that Turkish children who are exposed to everyday discourse that emphasizes the evidential base of beliefs may have a difficult time fully appreciating that the question asks for a mere guess in the case of diverse beliefs. We believe this might explain why a higher percentage of Turkish children fail in the DB task as compared to their English-speaking counterparts. In Turkish, the "*think"* verb is more likely to be used to refer to beliefs formed on an evidential basis. However, the DB task in the ToM scale does not provide children with this kind of basis yet it uses the "*think"* verb in order to indicate the diversity between the protagonist's and the target child's beliefs. Using the "*think"* verb in the DB task can thus mislead Turkish children to look for some evidence for selecting either option. Our observations show that many children spontaneously express having found evidence (e.g., "I can see the tip of the cat's tail in the bushes"), even when no such evidence actually exists. Eventually these children may overestimate the truth value of their own beliefs (since

they believe these are evidence-based) and report their belief which they believe represent the true state of affairs in the world. An obvious modification would be to replace the "*think*" verb with the verb "*guess*". A less obvious modification that might make the task more meaningful to Turkish-speaking children would be to provide evidence for both options. Specifically, where the task asks the target child's belief about where the cat might be (i.e., in the bushes or under the car), we could provide perceptual evidence that warrants either options (i.e., placing paw marks that lead to both locations). This modification would not alter the underlying logic of the task as either option would still be equally possible. However, children would have evidence on which to base both their own and the protagonist's beliefs. If the verb "*think*" is used more frequently in contexts where the evidence for the belief if explicit, using this verb should no longer hamper young children's performance in the DB task presented with evidence.

The lack of evidential base might cause Turkish children's poor performance on the EFB task, too. Wellman and Liu's (2004) EFB task is designed to be succinct. The vignette includes two possible options and the false belief of the protagonist with no justification for the reason for this (false) belief. Yet, in this task, the protagonist's false belief is expressed by the "*think"* verb. For children, who are accustomed to hearing the verb "*think*" in contexts where evidence is provided along with belief, the lack of justification might be impeding their performance. In other words, it might prevent children from understanding the belief of the protagonist and lead them to judge that the protagonist will (or must) search for the missing object in its real place despite the fact that this place is unknown by the protagonist. This explanation does not hold for children's relatively better but still poor performance in the CFB task

since this task justifies the verb "*think*" implicitly as the box used strongly implies its contents by its appearance (i.e., crayon box implies crayons) which form the basis of the protagonist's belief. Wellman and Liu (2004) based their version of the EFB task on Wimmer and Perner's (1988) task. The earlier version of this task is couched in a longer narrative where the reasons for the protagonist's beliefs are explicitly detailed (e.g., character puts item in one place, leaves, another character changes the place for a reason). This kind of explicit justification for the reason of characters' false beliefs in lacking in the Wellman and Liu (2004) version of the task. It is difficult to foresee whether children would perform better or comparably in the narrativized version of this task. Given that appropriate justification of the false belief is provided in the task vignette, it is possible that children's performance would be better as compared to the abbreviated version of the task. The current study will include both the abbreviated version (i.e., EFB-Original) and the narrativized version (i.e., EFB-Narrativized) to explore the possible effects of this manipulation.

Another reason for Turkish children's poor performance in the EFB and CFB tasks may lie in the selection of the verb "*think*" as compared to its pragmatically more accurate counterpart "*falsely think*" to talk about false beliefs. In Turkish, there are two verbs that are used to indicate one's own and the others' false beliefs: "zannetmek/sanmak" which can be translated to English as "*falsely think"*. All of the Turkish studies reviewed have opted to use the neutral "*think*" verb in contrast to the more accurate "*falsely think*" verb. This choice is due to considerations of comparability with children from other cultures who do not have such a mental state verb in their language. However, this practice may be adversely affecting Turkish children's performance who anticipate a true belief when they hear the verb "*think*"

and a false belief when they hear the verb "*falsely think*". Asking for a character's false belief with the "*think*" verb may actually be more cognitively demanding as it would require a more flexible and deeper understanding of mental state concepts as well as the ability to inhibit one's intuitions associated with everyday words. Replacing the "*think*" verb in the FB tasks with the verb "*falsely think*" is a viable alteration that could trigger a false belief context for Turkish children. Similar considerations were at play in Wellman and colleagues' (2006) study with Chinese preschoolers. These researchers opted to use the Chinese version of the verb "*falsely think*" in that study with Chinese preschoolers in order to eliminate the possible conflict between how the culture talks about false beliefs and how they are mentioned in FB tasks.

1.4 The Current Study

The current study takes a socio-cultural approach to the development of ToM abilities. This has two implications: (1) children's acquisition of mental concepts depend on their experiences (which also includes their linguistic experiences), (2) performance on tasks depend on how accurately these tasks reflect the real-world situations in which the linguistic markers for mental states are used, i.e. pragmatics.

If as proposed, Turkish speakers use the verb "*think*" most frequently in contexts where the reason for a particular belief is evident, then it is possible that using this verb when no reason is evident adversely affects children's performance. Additionally, if Turkish speakers mark the falsity of beliefs explicitly with the verb "*falsely think*" and especially in situations where there is inadequate evidential

support for a belief, then using the "*think*" verb to inquire about false beliefs would also hamper children's performance.

In the current study we aimed to: (a) test whether our intuitions about the appropriateness of different cognitive mental state verbs (i.e., "*guess, think, falsely think"*) are held by Turkish-speaking adults, (b) whether Turkish speaking adults' intuitions about the most appropriate word for each context would reflect the optimal conditions for best performance, and (c) whether 3-, 4-, and 5- year-old Turkish speaking children would show differences in their performance and differences in the sequence with which they pass ToM tasks based on the MSV used.

Studies 1 and 2 (with adults) were more exploratory in nature. For Study 1, as Turkish native speakers, we had certain intuitions about the conditions in which individuals opt for one MSV over another. We expected that "*guess"* would be deemed more appropriate than "*think"* in the diverse belief vignettes where there was equal likelihood for either of two options. We had weaker expectations for the verb "*falsely think*" in false belief vignettes/target questions (i.e., CFB, EFB). We expected that individuals may show differential performance based on the amount of evidence presented for a false belief. Our predictions for Study 2 paralleled our expectations from Study 1. Specifically, we expected that the verb "*falsely think"* may result in better performance (i.e., faster and accurate). However, we also thought there could be differences based on the type of FB task. The lack of previous research with this manipulation did not allow to make more specific hypothesis.

For Study 3, we generally hypothesized that the modified versions of the tasks would yield better performance with Turkish speaking preschoolers. Specifically, we expected better performance in the DB task with "*guess*" and "evidence" as compared to Wellman and Liu's (2004) original task. Similarly, we expected that children would perform better in the modified versions of the EFB task (i.e., EFB-*Falsely Think*, EFB-Narrativized) as compared to the original version of the task. In the same vein, we expected that children would perform better in the modified version of the CFB task (CFB-*Falsely Think*) as compared to the original version. In addition, we expected the modified versions of the tasks to reveal better correlations with known cognitive predictors of ToM (i.e., language and executive functions). Finally, we hypothesized that the sequence of ToM tasks would change if the scaling analysis are reconducted with the modified versions of the tasks which may consequently result in a more valid ToM scale for Turkish children.

CHAPTER 2

STUDY 1

This study investigated Turkish speaking adults' appropriateness judgements of mental state words in ToM tasks commonly used with preschool children. The study was carried out through an online survey, where adults read questions with blanks for them to fill with one of two mental state words provided in the options. The participants were fully briefed in the beginning of the study about the purpose (i.e., learning native speakers' judgments of appropriateness for mental state words). They were also told that the researchers were going to use these tasks with young children and aimed to ensure pragmatic and conversational appropriateness of these test questions for young children.

2.1 Method

2.1.1 Participants

One-hundred and fifty Turkish-speaking adults completed the online survey created through Qualtrics. Thirty-three of the participants were males and 112 of them were females; 5 participants did not state gender. Participants' age ranged from 18 to 61 years, M = 27.18, SD = 8.12. Participants' education ranged from elementary school to PhD education. Overall education level was high with 62% having a university a university degree and an additional 27% continuing their university education.

Participants were recruited through announcements made on social media websites (e.g., Facebook, Instagram). A snowball sampling method was used, where the initial announcement was digitally shared by the experimenter with friends and followers who were also encouraged to share the announcement with their own networks. This procedure allowed the survey to reach participants who were from diverse age and education groups. All participants agreed to participate by clicking on "I agree" button under the consent form that was seen on the first page of the survey. The ones who did not agree were directed to the end of the survey without seeing any of the questions.

2.1.2 Procedure

Participants were recruited through sharing Qualtrics links on social media websites such as Facebook and Instagram. Along with the link, they saw an invitation sentence and a very brief explanation of the study. When they clicked on the link, participants were first presented with the consent form. They were allowed to see the questions only after they agreed to participate. After the consent form, participants were informed about the purpose of the study, i.e., to provide a culture-specific baseline for comparison with child-samples, and they were reminded that there was no right or wrong answer and the study is conducted to get their opinion about the appropriateness of different mental state verbs in specific vignettes. In each survey vignette, they read a story about a protagonist but the word indicating the mental state of the protagonist was left blank. Participants were asked to select the most appropriate verb (among two choices) for the protagonist in the given situation. The survey consisted of 4 vignettes indicating diverse belief situations and 6 vignettes indicating false-belief situations (2 vignettes for each: explicit false belief,

narrativized explicit false belief, content false belief). The order of the questions was randomized for each participant by Qualtrics. After the participants completed the vignettes, they were asked to complete the demographic form (see Appendix A). The participants were able to pass any question that they did not want to answer. At the end, they were debriefed about the aim of the study more extensively. Qualtrics recorded their answers automatically. Answers were downloaded and edited by the experimenter herself and no other unauthorized person was able to reach the answers. The whole procedure and all materials were approved by Bilkent Ethics Committee. Survey took approximately 10-15 minutes to complete.

2.1.3 Materials

2.1.3.1 Demographics form

This form included questions about participants' age, native language, gender, and education (see Appendix A). It also asked the name of the participants, but that question aimed to avoid any confusion that might stem from filling the online questionnaire twice by mistake. Once this possible confusion was eliminated, participants were given alphanumeric ID numbers and their names were not included in the data set or any write up of the study.

2.1.3.2 Survey questions

In order to investigate Turkish adults' judgments about the appropriateness of different mental state words that could be used interchangeably (i.e., use of the verbs "*guess*" vs. "*think*"; "*think*" vs. "*falsely think*"), an online survey was created through Qualtrics. Participants were informed that the researchers were interested in the most

appropriate wording of several vignettes that would be used with preschool children. The participants were informed that task difficulty was not an issue, but the wording had to be clear and appropriate in accordance with how adults would talk to young children. The participants were then instructed to choose one of two choices of mental state words for each vignette to fill in the blanks where mental states would be. The survey included 4 vignettes that could be answered with a choice between the word "*guess"* or the word "*think"* in diverse belief situations. An additional 6 vignettes were presented to see whether participants would choose "*falsely think"* instead of "*think"* in false-belief situations (2 questions for EFB, 2 questions for narrativized EFB, and 2 questions for CFB). For each category, except narrativized EFB category in which one scenario was inspired from Perner and Wimmer (1988), one scenario was obtained from Wellman and Liu's (2004) scale while the remaining scenarios were created by the experimenter to closely resemble the original ones (Wellman and Liu's scenarios). The problem in the vignette remained the same (i.e., where a lost object or animal could be). Each task was supported by the images of the objects or the protagonists mentioned in the stories in order to maintain the similarity between the ways in which tasks are presented to children and to adults (See Appendix B for survey questions).

2.2 Results

The aim of the survey was to investigate whether Turkish adults prefer to use "*guess"* and "*falsely think"* instead of "*think"* in diverse belief and false belief vignettes, respectively. For this purpose, we conducted separate binomial tests for all vignettes, and we looked whether one of the MSVs provided in its choices ("*think" vs. "guess" or* "*think"* vs. "*falsely think"*) was selected significantly more than the other one.

Binomial tests investigated whether participants chose one of the options significantly more than chance (50%).

2.2.1 Diverse Belief

Four separate binomial tests revealed that there were significant differences in the proportion of the "*guess"* verb selected for the diverse belief vignettes (Vignette 1: 79.3%, Vignette 2: 76%, Vignette 3: 81.3%, and Vignette 4: 63.3%) as compared to the proportion of the "*think"* verb, *p*s < .001 for Vignettes 1, 2, and 3; $p = .001$ for Vignette 4. These results suggest that Turkish-speaking adults prefer the "*guess"* verb over "*think"* when there is no persuasive evidence indicating the truth of one of the beliefs.

2.2.2 Explicit False Belief (EFB)

Two separate binomial tests were conducted for two EFB vignettes. Analyses revealed that the proportion of the "*falsely think"* verb (71.3% and 67.3%) preferred for EFB vignettes was statistically higher than the "*think"* verb, *p*s < .001 for both vignettes. This finding indicates that Turkish speaking adults prefer to use the Turkish-specific "*falsely think"* verb more than the "*think"* when there is an explicit indicator for the falseness of one's belief and no justification was provided for this false belief.

2.2.3 Narrativized Explicit False Belief (Narrativized EFB)

Two other binomial tests were conducted to see whether Turkish-speaking adults were more likely to use the "*falsely think"* verb when there was a narrative scaffold

that explained the protagonist's false belief. However, the analyses indicated that the proportion of the "*falsely think"* verb (51.3% and 57.3 %) preferred in narrativized EFB conditions did not differ from the proportion of the "*think"* verb significantly; $p = .81$ and $p = .09$ for Vignettes 7 and 8, respectively. Non-significant results indicate that Turkish adults do not have a specific preference in terms of using "*falsely think"* or "*think"* when there is a story that is supporting one's false-belief.

2.2.4 Content False Belief

The last two binomial tests were conducted to see Turkish adults' preferences in terms of using "*falsely think"* vs. "*think"* in content false belief situations. Binomial tests showed that the "*think"* verb (60% and 70%) was preferred more than the "*falsely think"* verb in CFB vignettes; $p = .02$ and $p < .001$ for Vignettes 9 and 10, respectively. These results suggest that Turkish adults prefer to use the "*think"* verb when a past evidence (i.e., usual content of the container) is supporting one's false belief.

2.2.5 Summary of Results and Discussion

The results of the Study 1 informed us about Turkish speaking adults' judgments of the appropriateness of mental state verbs used in ToM tasks. Adults' preferences showed that some of the belief-related ToM tasks were not suitable for interchangeable use of Turkish mental state verbs (e.g., "*think"* vs. "*guess"*). According to the results of the binomial tests, Turkish speaking adults judged the "*guess"* (tahmin etmek) verb as more appropriate than the "*think"* (düşünmek) verb for situations in which the truth value of the belief was unknown to them. Similarly,

Turkish adults preferred to use the Turkish-specific "*falsely think"* (zannetmek/sanmak) verb over the "*think"* (düşünmek) verb for situations in which a belief state is presented explicitly as false without its reason. Once the falsity of one's belief was grounded on a reason, (e.g., by telling a narrative explicating the false belief as in the Narrativized EFB tasks), Turkish adults did not express a specific preference over "*falsely think"* vs. "*think"* verbs. We believe that null results might have stemmed from adults' differential appraisements of the narrativized bases provided for the protagonists' beliefs. Some of the participants might have judged that the narrative base was enough for the protagonist to form a true belief about what is in his/her mind; therefore, they might have gone with the "*think"* option. However, according to other participants, the narrative may not have sufficed to form an evidential base for the false belief; therefore, they might have judged the "*falsely think"* verb as more appropriate. This logic is also able to explain why participants preferred the "*think"* verb in CFB tasks significantly more than "*falsely think"* even though the task involves a false belief. CFB tasks are based on the idea that one person (e.g., the protagonist) has past evidence regarding the content of a descript container. According to aforementioned logic, past evidence regarding the content of a box and the descriptive features on it (e.g., pictures of the candy) would have sufficed to form a firm evidential base for one's belief and it warrants the use of the "*think"* verb given the pragmatics of Turkish language.

Study 1 provided us with the explicit appropriateness judgements of Turkish adults regarding the use of different mental state verbs in ToM tasks. However, we were also interested in studying whether using the verbs deemed more appropriate affected Turkish adults' performances on those tasks. To this end, we conducted a second

study with a new set of Turkish adults to investigate whether Turkish speaking adults who received these tasks with one MSV or the other (i.e., *"to think"* "düşünmek" *vs. "to falsely think"* "sanmak/zannetmek") would show differences in their reaction times to provide correct answers.

CHAPTER 3

STUDY 2

This study was designed to investigate whether Turkish speaking adults would show differences in their performance based on whether the false belief question included "*falsely think*" (sanmak/zannetmek) or "*think*" (düşünmek). While either verb can be used interchangeably in false belief contexts, it is possible that one of these words elicits faster and more accurate reactions. This experiment followed up on Study 1 where we asked adults their preference regarding either of these verbs. Study 1 showed that participants showed a preference for "*falsely think*" for EFB vignettes, no preference in narrativized EFB vignettes and a preference for "*think*" in CFB vignettes. Adults' explicit judgments provide insight into their intuitions about their judgments of appropriateness of mental state verbs in false belief contexts. However, whether their actual performance differs based on mental state verb would provide a deeper understanding of the effect of these verbs on adults' processing in false belief contexts.

3.1 Method

3.1.1 Participants

Sixty-one (64% female) Turkish-speaking adults participated to response time paradigm that was created through E-Prime® and that took place in Bilkent University Psychology Labs. All participants, except 2 graduate students, were undergraduates of

Bilkent University. Participants' age ranged from 18 to 24 years, *M* = 19.84, *SD* = 1.34. Participants were recruited through Bilkent University's e-mail system (BAIS). Although all participants made appointments voluntarily via e-mail, they were also asked to read and accept the consent form at the beginning of the experiment. Participants were given 10 GE 250/251 points for their participation.

3.1.2 Procedure

Participants were recruited through Bilkent University's e-mail system (BAIS). In the mail, they were briefly informed about the study, and they were asked to make appointments if they wished to participate. Once they made an appointment with the experimenter through e-mail, they arrived in Bilkent University Psychology Labs at their scheduled times and they were informed about what they need to do during the test. Then, the experimenter started the testing session on computer, made sure that the consent form was on the screen and the participant was wearing the headset, and she left the testing room. All participants agreed to participate after they read the consent form. After they approved the consent form, they saw the written protocol. Once they were done with reading the protocol, they pushed "Enter" button and testing started.

First, they saw a question screen which included the pictures of protagonists and objects/locations mentioned in the vignettes and target questions. While this screen was on, a pre-recorded voice presented the false belief vignette and asked the target question (see Appendix C for questions). As soon as question was asked, the screen with the pictures disappeared, and another screen with answers appeared. The second screen did not contain any pictures, instead the two possible answers were presented

in writing. No auditory stimulus was given at this stage. Participants had to read the options and chose one of the presented words by pushing V or N button on keyboard (V for the word on left side and N for the word on right side). Once they answered a question, the other question appeared on screen automatically and participants were not able to go back and change their answer for previous questions. The order of questions was randomized by E-Prime®.

When all questions were answered, participants saw the debriefing information on the screen, and they informed the experimenter who was waiting outside of the testing room. The experimenter asked participants whether any technical problem occurred during testing (none of the participants reported such problem) and handed the demographic form. Participants filled out the demographic form and left the testing room. Their answers (in terms of both accuracy and response time) were recorded by the program automatically. Participants' ID numbers were also written on demographic forms in order to match the data on computer and on demographic forms. The whole procedure and all materials were approved by Bilkent Ethics Committee. Participants who were enrolled in GE 250/251 courses (non-credit, activity courses) were given 10 GE 250/251 points for their participation.

3.1.3 Materials

3.1.3.1 Demographics form

The demographics form was the same with the one that was used in Study 1 (see Appendix A), except with a small change: no identifying information was asked since

there was not a risk of participating more than once. Demographic forms were given as printed forms after participants completed the experiment.

3.1.3.2 False belief stimuli

Participants answered 6 false-belief questions embedded in 6 false-belief tasks on E-Prime® . The false-belief tasks used in this study were very similar to what is normally used with preschool children, both in terms of content and presentation. In order to maintain high similarity, the false belief vignettes and the following false belief question for each of the 6 tasks were provided auditorily. The participants saw pictures of protagonists and objects while listening to the auditory presentation of the vignette and the question. After the participants listened to the vignette and the target question, the screen with the picture of the protagonists and objects was replaced with a screen with the two options in writing. The participants were expected to read the possible choices. (see Figure 1 for sample screens)

Two content false belief (CFB) tasks were used in which participants' performance in response to "*think"* vs. "*falsely think"* verbs used in target questions were measured. Two explicit false-belief (EFB) and 2 narrativized EFB tasks measured whether participants' performance would differ in response to the use of "*think*" vs. "*falsely think*" in false-belief vignettes. One CFB task and one EFB task were taken from Wellman and Liu's (2004) ToM battery and one narrativized EFB task was inspired from Perner and Wimmer's (1988) unexpected change task. For each of these tasks (CFB, EFB, or narrativized EFB), another similar task was created by the experimenter so that there would be a pair of tasks targeting the same kind of false belief understanding (see Appendix C for the tasks). However, two tasks of the same

kind differed from each other in terms of the mental state verb (MSV) used in the target question or vignette. In other words, one tasks' target question/vignette included the "*think"* verb while the target question/vignette of the other task in the same pair included the verb "*falsely think"*. Within a pair, which question was asked with "*think*/*falsely think"* verb was counterbalanced across participants.

As it can be seen in Figure 1, during the auditory presentation of CFB vignettes and target questions, the picture of the descript container was always presented on the left of the computer screen while the picture of the protagonist was always located on the right side. On the second screen where participants saw the answers in writing, though, the presentation order of the answers was counterbalanced between participants. The right answer of the CFB task (i.e., usual content of the descript box) was presented on the left of the screen for half of the participants and on the right for the rest of them. However, we should note that the auditory presentation of the possible contents in target questions and the visual presentation of the corresponding words in the second screen always followed the same order, assuming participants read the words from left to right. For example, if the target question was asked as "*What does Canan falsely thinks that is inside the box? Candy or toy?*", writing for "candy" appeared on the left of the screen and the writing for "toy" appeared on the right. This counterbalancing ensured that the right answer for CFB tasks did not appear on the same side of the screen all the time.

During the auditory presentation of EFB/narrativized EFB vignettes and target questions, though, the picture of the protagonist was always presented on the middle of the screen while the picture of the 2 locations where the object might be were

presented on the left and the right sides of the screen, under the protagonist's picture. To all participants, the pictures of the two possible locations for an object were presented in the same order on the first screen (i.e., real location on the left and the location in protagonist's mind on the right) due to the structure of false belief questions in our study: they always tell the real location of an object and then where protagonist thinks/falsely thinks the objects is. However, similarly to CFB tasks, the presentation order of the answers was counterbalanced between participants on the second screen where participants saw the answers in writing. Specifically, the word corresponding to the right answer of an EFB task (i.e., the location in protagonist's mind) was presented on the left of the screen for half of the participants and on the right for the rest of them. Again, the auditory presentation of the possible locations in target question and the visual presentation of the corresponding words in the second screen always followed the same order, assuming the participants read the words from left to right. For example, if the target question was asked as "*Where will Elif look to find her gloves? The cupboard or her backpack?",* writing for "cupboard" appeared on the left of the second screen and the writing for "backpack" appeared on the right. This counterbalancing also aimed to prevent participants from getting the right answer just because they got used to answer appearing on the same side of the screen. If the words' presentation order on the second screen was not counterbalanced between participants and the words were always presented in the same order with the pictures of possible locations (in the first screen), participants would have seen the right answer on the right side of the screen and given accurate answers to all questions by pushing the same button on keyboard (N).

Task Type	First Screen	Second Screen	
CFB	 *"This is a candy box, but it actually contains a toy. Canan has never seen what is inside the box. What does Canan falsely think that is inside the box? Toy or candy?"*	Toy V	Candy N
EFB	   *"This is Elif. Elif searches her gloves. Elif's gloves are either in her backpack or in the cupboard. Actually, her gloves are really in the backpack, but she falsely thinks that her gloves are in the cupboard. Where will Elif look to find her gloves? The cupboard or her backpack?"*	Cupboard V	Backpack N

Note. Auditory material italicized.

Figure 1. Sample Screens from Study 2

3.2 Results

In this study, we aimed to investigate whether the response times of Turkish speaking adults when answering false belief (FB) questions would depend on the MSVs ("*think"* or "*falsely think")* that were used in the false-belief vignettes or questions. Specifically, we hypothesized that Turkish speakers would reply more rapidly if the FB tasks were presented with the Turkish-specific "*falsely think"* verb. A quick glance at the data showed that adults were not accurate 100% of the time. Given the variability in accuracy (Table 3), we decided to analyze both accuracy and response time in relation to mental state verb (MSV) used in false belief tasks. Since the accuracy data was categorical in nature (pass/fail) and since all participants answered both versions of one kind of FB task, we used McNemar's chi-square tests in order to compare participants' accuracy rates in two versions (i.e., with "*think"* vs. "*falsely think"* verb) of a task. When analyzing the response time in relation to MSV used in the false belief vignette/target question, we conducted paired-sample t-tests and compared the average response times for the versions presented either with "*think"* vs. "*falsely think"*. Before each t-test, we eliminated the participants who failed at either of the related tasks and the outliers whose response time was 3 SD from the mean.

Table 3. Number of Adult Participants Who Failed FB Tasks in Study 2

Total # of participants who:	Failed 1 Task	Failed 2 Tasks	Failed 3 Tasks	Failed No Task
	8	10	4	39

Table 3 (cont'd)

# of mistakes in each task:	CFB *Think*	CFB *Falsely Think*	EFB *Think*	EFB *Falsely Think*	EFB-Narrativized *Think*	EFB-Narrativized *Falsely Think*
	9	6	9	6	8	2

Note. None of the participants failed more than 3 tasks.

3.2.1 CFB-Accuracy

McNemar's test revealed that using "*think"* vs. "*falsely think"* in the target questions of CFB tasks did not affect participants' success on those tasks, $p = .55$. It means Turkish adults' success on CFB tasks is not affected by verb use.

3.2.2 CFB-Response Time

We firstly eliminated participants who failed at CFB tasks (n = 13) and outliers (ID # = 108 & 131) whose response times differed 3 SD away from the mean response time for CFB tasks (n = 2). It resulted in 46 participants. A paired-sample t-test revealed that using "*think"* ($M = 1336.41$, $SD = 700.68$) vs. "*falsely think"* ($M = 1240.04$, $SD = 601.89$) in CFB tasks did not affect the participants' response time, $t\ (45) = -.80$, $p = .43$. It means that Turkish adults' speed of responding to CFB questions is not affected by verb use.

3.2.3 EFB-Accuracy

McNemar's test revealed that using "*think"* vs. "*falsely think"* in the vignettes of EFB tasks did not affect participants' success on those tasks, $p = .38$. It means Turkish adults' success on EFB tasks is not affected by verb use.

3.2.4 EFB- Response Time

Before conducting t-test, participants who answered any EFB question wrong (N = 10) were eliminated from the data. Also, one participant (ID # = 141) whose response time was 3 SD away from the average response time of EFB tasks was eliminated. It resulted in 50 participants. A paired-sample t-test revealed that using "*think"* (*M* = 840.32, *SD* = 495.06) vs. "*falsely think"* (*M* = 834.16, *SD* = 399.45) in EFB tasks did not affect the participants' response time, *t* (49) = -.10, *p* = .92. It means that Turkish adults' speed of responding to EFB questions is not affected by the verb used in the EFB vignette.

3.2.5 Narrativized EFB- Accuracy

McNemar's test revealed that using "*think"* vs. "*falsely think"* in the vignettes of narrativized EFB tasks did not affect participants' success on those tasks, *p* = .11. It means that Turkish adults' success on EFB tasks which were supported by narrative bases is not affected by verb use.

3.2.6 Narrativized EFB- Response Time

Like the all response time analyses conducted so far, participants who answered any narrativized EFB question wrong (N = 10) were eliminated from the data. No outlier was observed. It resulted in 51 participants. A paired-sample t-test revealed that using "*think"* (*M* = 912.90, *SD* = 476.96) vs. "*falsely think"* (*M* = 988.43, *SD* = 455.52) in narrativized EFB tasks did not affect the participants' response time, *t* (50) = .93, *p* = .36. It means Turkish adults' speed of responding to narrativized EFB questions is not affected by the verb used in the vignette.

3.2.7 All FB Tasks

In Study 2, all participants responded to 6 FB tasks. Three FB tasks were presented with the "*think*" verb (1 CFB, 1 EFB, and 1 Narrativized-EFB task) whereas the other 3 (1 CFB, 1 EFB, and 1 Narrativized-EFB task) were presented with the "*falsely think*" verb. Although participants' performances in two versions of the same task (e.g., CFB with "*think"* vs. CFB with "*falsely think"*) did not differ from each other, we thought their overall performance (when all tasks combined) would be affected by verb use. We calculated the overall performance by aggregating participants' performances (accuracy or response time) under two groups: tasks presented with the "*think"* vs. "*falsely think"* verbs. Then we compared the averages of the two groups in terms of accuracy or response time.

3.2.7.1 Accuracy

A paired-sample t-test revealed that using the verb "*falsely think"* (*M* = 2.77, *SD* = .42) versus *"think"* (*M* = 2.57, *SD* = 0.69) facilitated Turkish adults' accuracy in FB tasks, *t* (60) = -2.56, *p* = .01.

3.2.7.2 Response time

First, participants who answered any FB task wrong (N = 22) were eliminated from the data. No outlier was observed. It resulted in 39 participants. A paired-sample t-test revealed that using the verb "*think"* (*M* = 3100.49, *SD* = 1214.38) vs. "*falsely think"* (*M* = 2924.85, *SD* = 1036.66) in FB tasks did not affect the participants' response time, *t* (38) = -.92, *p* = .36.

3.2.8 Summary of Results and Discussion

The results of Study 2 informed us about whether Turkish adults' processing speed in false belief tasks would depend on the MSVs used in these tasks ("*think"* vs. "*falsely think")*. Paralleling the predictions made for Study 1, we expected Turkish adults to provide the correct answer faster for false belief tasks when the vignette or the target question of the task included the verb "*falsely think*", instead of the verb "*think*". When the different versions of each task were compared with each other (with "*think"* vs. "*falsely think"*), no performance differences were observed. Although Study 1 had shown that Turkish adults preferred the verb "*think"* in CFB tasks and the verb "*falsely think"* in EFB tasks, neither the accuracy rate nor the response time varied by the MSV used in these tasks. Using "*think"* vs. "*falsely think"* in the Narrativized EFB tasks did not affect adults' performances on these tasks for which they had not expressed a MSV preference in Study 1. When adults' performances were evaluated out of 6 FB tasks, however, the accuracy rate in FB tasks differed by verb use. Specifically, participants were more accurate when the FB tasks were presented with Turkish-specific "*falsely think*" verb. The response time was not affected by the verb used in the FB tasks. Collectively these results show that while for adults who respond accurately to FB question, the type of MSV used does not make a difference. However, adults are more likely to respond more inaccurately when the task includes the "*think*" version of the verb as compared to the "*falsely think*" version. The speakers of a language may not use MSVs 100% correct all the time, yet they can convey or understand the true message with the help of the context. However, we believe that children who have limited experience for the misuse of MSVs would be more sensitive to interchangeable use of these verbs: they might benefit more from the proper usage of MSVs while suffering more from the misuse. In order to test this

idea, we designed a third study with Turkish children. We investigated whether Turkish-speaking children's performance in belief-related tasks of ToM battery (DB and FB tasks, Wellman and Liu, 2004) was affected by the verb used in these tasks.

CHAPTER 4

STUDY 3

Prior research on ToM scaling provided us with children's success in each ToM task. If we put the percentages from multiple studies together, what we see is that Turkish-speaking children are less likely to pass belief-related tasks (DB, EFB, and CFB) of Wellman and Liu's (2004) ToM battery than their peers from other countries such as the US, Australia, and China (e.g., Etel & Yağmurlu, 2015; Selçuk, Brink, Ekerim, & Wellman, 2018; Ilgaz, Allen, & Haskaraca, 2019; Shahaeian et al., 2011; Wellman et al., 2006; Wellman & Liu, 2004). The current study was conducted in order to investigate whether Turkish-preschoolers' poorer performance in belief-related tasks could stem from the fact that these tasks were originally created for English speaking children and they were used in their directly-translated format in the aforementioned Turkish studies. We believe, direct translations of these tasks are not sensitive to the Turkish-specific, pragmatic and semantic nuances between mental state words (e.g., "*think"* vs. "*guess"* and "*think"* vs. "*falsely think"*). We modified Wellman and Liu's belief-related tasks so that they would be more in-line with Turkish's pragmatics and semantics. We designed a within-subject study in which all participants were tested on every possible version of the tasks over two sessions to investigate if their performances differed across the different versions of the same tasks. We also measured language and executive function abilities of children since these abilities

have consistently been found to be related to ToM abilities in different cultures that have been studied (see for reviews, Astington & Jenkins, 1999; Hughes, 1998; Miller, 2006; Milligan, Astington, & Dack, 2007; Moses & Tahiroğlu, 2010). We controlled whether these abilities correlated with children's total ToM abilities in our study as well and we also looked whether they played a role while children's performances fluctuated between the modified versions of one type of task (DB, EFB, or CFB).

4.1 Method

4.1.1 Participants

Sixty Turkish-speaking preschoolers (30 girls) participated in the study. Children and their parents were residents of Manisa, and all children were attending preschool. Children's age ranged from 3.1 to 5.9 years (M = 4.6, SD = .89). Their mothers were between 24 and 55 years of age (M = 32.8, SD = 5.7) and the fathers were between 24 and 54 (M = 36.3, SD = 6.0). Both mothers' and fathers' education ranged from elementary school to PhD; however, fathers were more likely to work in full-time jobs (93%) than mothers (52%). Income reports of the families are as follows: 6 families were earning under 1650 TL/Month; 16 families were earning between 1650-3500 TL/Month; 19 families were earning between 3500-5000 TL/Month; 9 families were earning between 5000-7000 TL/Month; and, 4 families were earning more than 7000 TL/Month. These numbers should be approached carefully since they may not be reflecting the true income of the families: many fathers who filled out the demographic form reported that they only considered their own income when answering income-related questions in the form and did not include their wives' earnings even though their wives were working or they had an alternative income

stream (e.g., money inherited from the family). Based on the recruitment strategy (i.e., data was collected in private preschools whose monthly tuitions ranged between 750 TL to 1200 TL), it is possible to say that children in this study were coming from middle socioeconomic status (SES) backgrounds.

Most of the preschoolers (N = 56) were recruited through visits to preschools and through announcements that were sent to families by the preschool managements. Those children were tested in a private room in their preschools. The rest of the participants (N = 4) were acquaintances of the experimenter. Although all 4 of these children were attending preschools, they were tested at their homes since their preschools were not included in the school list approved by the Ministry of Education. Children participated in the study only after both of their parents had signed the consent form. Children's verbal assent was obtained at the beginning of each testing session. All materials and procedures were approved by Bilkent University Ethics Committee and Ministry of Education. Upon the completion of the tasks, children received stickers.

4.1.2 Materials

4.1.2.1 Demographic form

This form included questions about parents' ages, education levels and occupations; number of siblings and their birthdates; and monthly household income and how parents perceive their income levels (See Appendix D).

4.1.2.2 Theory of Mind battery of Wellman & Liu (2004) and modified versions of belief-related tasks.

4.1.2.2.1 Diverse desires (DD)

This task was devised to assess children's ability to judge that two people (e.g., the child himself/herself and the protagonist) may have different desires about the same objects. In this task, the child is shown the picture of the protagonist along with two other pictures one of a cookie and the other of a carrot. Then the child is told "*This is Duygu/Osman. Duygu/Osman is hungry, and s/he wants to eat something. Here are two things that s/he can eat: a carrot and a cookie. Which of those would you like best, carrot or cookie?*" This question is asked to determine child's own desire. Once the child answers own-desire question (e.g., *cookie*), experimenter tells that protagonist's favorite food is the other one, i.e., the one that is not selected by the child: "*It is a good choice, but Duygu/Osman likes the carrot best, s/he does not like cookies, what s/he likes best is carrot.*" Then the experimenter asks the target question: "*It is time to eat now. Duygu/Osman can choose just one of these foods. Which one will s/he choose, carrot or cookie?*" A child is considered as successful if s/he answers the target question in line with protagonist's desire, i.e., *carrot*, in this example.

4.1.2.2.2 Diverse beliefs (DB-Original)

This task is supposed to assess children's ability to judge that two people (e.g., the child himself/herself and the protagonist) may have different beliefs regarding the location of the same object where the true location is unknown. The child is shown the picture of the protagonist along with two other pictures one of a set of bushes and the other of a parked car. Then the child is told "*This is Dilek/Tolga. Dilek/Tolga wants to find her/his dog. Her/his dog is either in the bushes or under the car. Where do you think the dog is, in the bushes or under the car?*" This question is asked to

determine child's own belief. After the child answers the own-belief question (e.g., *in the bushes*), experimenter says the protagonist thinks that the dog is in the other place, i.e., the one that is not selected by the child: "*It is a good idea, but Dilek/Tolga thinks that the dog is under the car.*" Then the experimenter asks the target question: "*Where will Dilek/Tolga look for her/his dog, in the bushes or under the car?*" A child is considered as successful if s/he answers the target question in line with protagonist's belief, i.e., *under the car*, in this example.

4.1.2.2.3 Diverse beliefs with "guess" (DB-*Guess*)

As Study 1 revealed, Turkish adults prefer to use the "*guess"* verb over the "*think"* verb when there are two different and possible beliefs about an object or situation, and where the true state of affairs is unknown. To test whether adults' intuitions about the appropriateness of the "*guess"* verb in these situations is in line with preschoolers' preferences we created a pragmatically-modified version of the DB task. The only difference between this task and its original version is the use of the verb "*guess"* in place of the verb "*think".* In this task, the child first saw the picture of the protagonist along with two other pictures one of a tree hole and the other of a hole in the ground. Then, experimenter says "*This is Ayşe/Ali. Ayşe/Ali wants to find her/his rabbit. Her/his rabbit is either in the tree hole or in the hole in the ground. Where do you guess the rabbit is, in the tree hole or in the hole in the ground?*". This question is asked to determine child's own guess. After the child answered the question (e.g., *in the tree hole*), experimenter told that the protagonist guesses the rabbit is in the place not selected by the child. This was followed by the target question: "*It is a good guess, but Ayşe/Ali guesses that the rabbit is in the hollow* (option not selected by the child). *Where will Ayşe/Ali look for her/his rabbit, in the tree hole or in the hollow?*"

A child was considered as successful if s/he answered the target question in line with protagonist's guess.

4.1.2.2.4 Diverse beliefs with evidential clues (DB-Evidence)

As is the case with both original and pragmatically-modified versions of the DB tasks, DB-Evidence task is also about a missing pet (a cat); however, this time, both pictures depicting cat's possible hiding places (e.g., a box or under the sofa) have the same evidential clue, i.e., paw marks, leading to each of these places. The addition of these clues (paw marks) does not aim to give advantage to one belief over the other; the aim is to provide the evidential base that is likely to be taken as reference for either of these locations. This modification was done to allow Turkish speaking children to better evaluate each of these options warranting belief status (as opposed to a chance level guess). First, the child saw the picture of the protagonist along with two other pictures of one of a half-closed box and one of a sofa, and the identical paw marks leading to each of these objects: "*This is Ceren/Cem. Ceren/Cem wants to find her/his cat. Her/his cat is either in the box or under the sofa. Look, there are paw marks leading to both places. Where do you think the cat is, in the box or under the sofa?*" (Own-belief question) After the own-belief question was answered (e.g., *in the box*), experimenter asked a control question ("*Why do you think that?")* in order to make sure that the child reports the presence of the paw marks as basis for their own belief. If children did not refer to the paw marks, we would not with any certainty know whether the child takes the evidence into consideration. The absence of reference thus would show that the modified task was interpreted the same way as the original the DB task. All children in our data set either told that their belief is due to paw marks or they pointed to the paw marks in the pictures. Next, the experimenter stated that the

protagonist thinks the cat is in the other place which is not selected by the child: "*It is a good idea, but Ceren/Cem thinks that the cat is under the sofa.*" Then the target question is asked: "*Where will Ceren/Cem look for her/his dog, in the box or under the sofa?*" A child was considered as successful if s/he answered the target question in line with protagonist's belief, i.e., *under the sofa*, in this example.

4.1.2.2.5 Knowledge access (KA)

This task assesses whether children understand that perception leads to knowing. Specifically, whether they understand that someone who has seen inside a nondistinctive box will know its contents, but another person (e.g., the protagonist in the task) who did not see inside the box will be ignorant about its contents. For this task, child was shown a small, nondescript box. The experimenter stated "*Here is a box. What do you think is inside this box?*" Any answer including "I do not know" was deemed as acceptable. Then the box was opened by the experimenter and its content were revealed: "*Let's see what is inside the box. There is a small dog inside!*" Once it was ensured that child saw the content clearly (child is allowed to hold briefly and take a glance at the dog if s/he wants to), the experimenter closed the box and asked the memory-control question: "*Okay, tell me, what is inside the box?*" The inside of the box was visible to the child until s/he answered the memory-control question correctly. After this the box was closed and contents no longer visible. Following this, the protagonist was introduced, and the target and memory questions were asked, respectively: "*Now comes Merve/Barış. Merve/Barış has never seen inside this box. Does Merve/Barış know what is in this box? Did Merve/Barış see inside this box?"* A child was considered as successful if s/he answered both target and memory question correctly (i.e., "*no.*").

4.1.2.2.6 Explicit false belief (EFB-Original)

This task originally aims to understand if children are able to judge whether a person with a false belief will act in accordance with his belief rather than the true state of affairs in the world. In this task child was shown a picture of the protagonist and told that "*This is Elif/Emre. Elif/Emre is looking for her/his gloves. Her/his gloves are either in the drawer"* (experimenter puts the picture of a drawer in front of the child) *"or in the cupboard."* (Experimenter puts the picture of a cupboard in front of the child) *"In reality, Elif/Emre's gloves are in the drawer. But Elif/Emre thinks that her/his gloves are in the cupboard.*" Then target and control questions were asked to child in sequence: "*Where will Elif/Emre look for her/his gloves, in the drawer or in the cupboard?*" and "*Where are Elif/Emre's gloves really, in the drawer or in the cupboard?*" Child is expected to reason about protagonist's false belief (i.e., thinks that the gloves are in the cupboard) and predict that the protagonist will act in accordance with her/his false belief (i.e., look for the gloves in the cupboard) even though the true location of object is different and is known to the child. Therefore, the child needed to answer the target question "*in the cupboard*" and the reality question "*in the drawer*" to be considered successful in this task.

4.1.2.2.7 Explicit false belief with "falsely think" (EFB-*Falsely Think*)

This task is the pragmatically-modified version of EFB task and it is identical to the original task except the fact that it includes the verb "*falsely think"* (zannetmek/sanmak) in place of the verb "*think"* when indicating protagonist's false belief. The task started with introducing the protagonist and his/her search for an

object and continued with the presentation of false belief: "*This is Ecem/Ercan. Ecem/Ercan is looking for her/his crayons. Her/his crayons are either in the toy box"* (experimenter puts the picture of a toy box in front of the child) *"or in the backpack."* (the picture of the backpack is put in front of the child. *"Really, Ecem/Ercan's crayons are in the toy box. But Ecem/Ercan falsely thinks her/his crayons are in the backpack.*" Following this the experimenter asked the target and control questions in sequence: "*Where will Ecem/Ercan look for her/his crayons, in the toy box or in the backpack?*" and "*Where are Ecem/Ercan's crayons really, in the toy box or in the backpack?*" The child needed to answer the target question "*in the backpack*" and the control question "*in the toy box*" to be considered successful in this task.

4.1.2.2.8 Explicit false belief with a narrative base (EFB-Story)

This version of EFB tasks is a combination of two different false belief tasks: the original EFB task of Wellman and Liu's (2004) ToM scale and Perner and Wimmer's (1988) change of location task: inspired from change of location task, we provided a narrative-formed justification for the protagonist's mistaken belief in the original EFB task. Narrative thought includes protagonists' mental states and explicates the relations between their mental states and their actions. It is possible that children find false belief tasks that are in narrative format more meaningful since the reasons for characters' actions can be interpreted with the scaffold of the narrative. It is also more likely that, in Turkish, "*think"* verb (düşünmek) is used in order to refer beliefs that are based on some kind of evidential information. The narrativized justification can work as an evidential base for the protagonist's belief and legitimize the use of "*think"* verb; therefore, decrease the pragmatic insensitiveness of the task. While

adding the narrative, we also wished to keep the overall cognitive load and difficulty as comparable as possible to other EFB tasks by sticking to the format where the pictorial scaffold is uniform (character picture, 2 location) pictures. Therefore, we did not show 5 pictures depicting the whole story line as it is usually done with Perner and Wimmer's change of location task, instead, we used two pictures displaying the possible places for the object and a picture of the character.

The task began with the introduction of the protagonist and the possibility that the object the protagonist wishes to attain could be in either of two locations: "*This is Fatma/Fatih. Fatma/Fatih is looking for her/his chocolate. Her/his chocolate is either in the refrigerator"* (experimenter puts the picture of a refrigerator in front of the child) *"or in the kitchen closet"* (the picture of the closet is also put). Following this, the false belief of the protagonist and a narrativized basis for this false belief were presented to the child: "*In reality, Fatma/Fatih's chocolate is in the refrigerator. But Fatma/Fatih thinks her/his chocolate is in the kitchen closet because s/he left the chocolate box in the closet before s/he went to play outside. However, when s/he was outside, her/his mother took the chocolate box from the closet and put it in the refrigerator so that the chocolate would not melt. Then, while Fatma/Fatih was still playing outside, her/his mother went to her bedroom and took a nap."* Next, the following questions were asked: "*Where will Fatma/Fatih look for her/his chocolate when s/he comes back and wants to eat chocolate, in the refrigerator or in the kitchen closet?*" (Target question) and "*Where are Fatma/Fatih's chocolates really, in the refrigerator or in the kitchen closet?*" (Control question) The child had to answer the target question "*in the kitchen closet*" and the control question "*in the refrigerator*" to be considered successful in this task.

4.1.2.2.9 Content false belief (CFB- Original)

The aim of this task is to investigate whether a child is able to judge another person's (e.g., protagonist's) false belief about the content of a distinctive container when the child knows that the actual content is different than what is suggested by the appearance of the box, (e.g., a crayon in a candy box). In this task, the child was shown a distinctive box or drawer (e.g., a candy box) and asked "*Look, there is a box! What do you think is inside this box?"* All children in our sample were familiar with the Bonibon box we showed them and answered the question correctly (e.g., candy, Bonibon, chocolate, etc.). Once the child answered the question, the box was opened by the experimenter and its real content was revealed: "*Let's see what is inside. Mmm, it is really a crayon inside.*" Next the box was closed, and the experimenter asked the memory-control question: "*Okay, what is inside the box?*" If the child responded incorrectly, the content of the box was shown again, and the question repeated. Following this, the picture of the protagonist was shown, and the following questions were asked: "*Here comes Canan/Can. Canan/Can has never seen inside this box. What does Canan/Can think is in the box, a candy or a crayon?*" (Target question) "*Did Canan/Can see inside this box?*" (Control question). The child was expected to answer the target question based on the appearance of the box (i.e., "*candy*") and the control question as the protagonist has not seen inside the box (i.e., "*no*") in order to be successful in this task.

4.1.2.2.10 Content false belief with "falsely think" (CFB-*Falsely Think)*

The original and pragmatically-modified versions of CFB tasks were identical except that the modified version used the verb "*falsely think"* (zannetmek/sanmak*)* in place

of the verb "*think"*. In addition, the container used in this version was different than the one used in the original version (i.e., a crayon box with a scotch tape inside, instead of a candy box). In the task, children were shown a crayon box and asked: "*Look, there is a box! What do you think is inside this box?"* All children in our data set were familiar with the crayon box and they answered this question correctly. Next, the experimenter opened the box and displayed the real content of the box: "*Let's see what is inside. Mmm, it is really a role of scotch tape inside.*" Box is closed and memory-control question is asked: "*Okay, what is inside the box?*" Child is shown inside of box until s/he gets the memory-control question right. Then, the picture of the protagonist is introduced, and the questions are asked: "*Here comes Aslı/Ahmet. Aslı/Ahmet has never seen inside this box. What does Aslı/Ahmet falsely think is in the box, a crayon or a scotch tape?*" (Target question) "*Did Aslı/Ahmet see inside this box?*" (Control question) The child is expected to answer target question "*crayon*" and control question "*no*" in order to be considered as successful.

4.1.2.2.11 Hidden emotion (HE)

This task aims to reveal whether the child understands that a person (e.g., the protagonist) can feel one emotion while displaying a different, even an opposite, emotion. Initially, child is shown an three face drawings (a frowning, a neutral, and a smiling face) and is told that the experimenter is going to tell a story about a child in which the protagonist might be feeling sad (pointing to sad face), happy (pointing to happy face), or neutral (pointing to neutral face). Then the child is asked to identify each of the faces to check that s/he knows these emotions and emotional expressions. If the child fails to show the correct face for one of these emotions, the experimenter presents the faces again. If the child knows all expressions correctly, the task is

continued: "*Now I will tell you the story. Afterwards, I will ask you how this child really feels and which expression s/he is displaying on her/his face. How this child feels and the expression on her/his face might be same or different*." Then the picture of faces is put aside (still visible to the child), and the experimenter shows a picture of the protagonist whose face is blank (i.e., no eyes, no mouth, no nose) and tells the story: "*This story is about Melis/Mert. Melis/Mert's aunt is just back from a trip. Before she left for this trip, she promised to bring a toy for Melis/Mert, but instead she brought clothing instead of a toy. However, Melis/Mert does not like clothes, what s/he actually likes are toys. Yet, Melis/Mert has to hide her/his feelings since, if his/her aunt finds out her/his real feelings, she would never get a present for her/him.*" Then the child is asked two memory-control questions: "*What did Melis/Mert's aunt get for her/him?*" and "*What would happen if Melis/Mert's aunt finds out her/his real feelings?*" If the child fails to provide the true answers (e.g., "*dress* or *clothes*" and "*she would never buy anything for her/him*") for any of these questions, the story is told one more time. If the child answers memory questions correctly, target questions are asked: "*How did Melis/Mert feel when her/his aunt gave the cloth? Happy, sad, or neutral?*" and "*Which emotion did Melis/Mert display on the face when her/his aunt gave the cloth? Happy, sad, or neutral?*" A child is considered as successful if s/he answers both the hidden emotion (and the real emotion target questions correctly (i.e., "*sad*", "*happ*y", respectively).

4.1.2.2 Executive Function

Children's executive function (EF) abilities were measured by Dimensional Card Sorting Task (DCCS) devised by Zelazo (2006). Two types of cards are used in this task: one has the picture of a red car on it while the other one has a blue elephant.

Children are asked to sort these cards according to color or shape on them. The game begins with color rule and continues with shape rule. There is also a third rule (border rule) that requires switching between shape and color rules. According to Zelazo and Muller (2002), DCCS measures children's inhibition (i.e., not sorting according to old rule), and cognitive flexibility (i.e., switching between rules).

4.1.2.2.1 Color Game

DCCS starts with the color game in which children are asked to sort 8 cards (including 2 training cards) according to their colors (blue or red). Four cards are blue and 4 are red. The cards are presented in an order: at most 2 cards having the same color are presented consecutively. Training cards (one red and one blue card) are sorted by the experimenter in order to familiarize the children with the task. While children are sorting the cards, they are reminded of below rule before each card: *"Now we are playing the color game. In the color game, blue ones go here (pointing the box having a blue elephant on it) and reds go there (pointing the other box having a red car on it). This one is blue/red. Where does it go?"*

Each correct sorting gets 1 point, sorting all of the cards correctly results in 6 points. All children in our data set got full points (6) in color game.

4.1.2.2.2 Shape Game

After children sorted 8 cards (2 trial cards + 6 play cards) according to color rule, they are informed that "they are not playing color game anymore" and they are going to play a new game. For this task/game, children are told to sort the cards (6) according to shapes on them: elephant or car. Three cards have elephants and 3 cards have cars

on them. There is no training for shape game. The cards are presented as they are presented in color game, i.e., at most 2 cards having the same shape appeared consecutively. And again, a similar rule is reminded to children before each card: *"Now we are playing the shape game. In the shape game, cars go here (pointing the box that has a red car on it) and elephants go there (pointing the other box having a blue elephant on it). This one is car/elephant. Where does it go?"*

Each correct sorting gets 1 point, sorting all of the cards correctly results in 6 points. If a child successfully sorts at least 5 of the 6 cards in this game, s/he qualifies to play the more complicated border game.

4.1.2.2.3 Border Game

In the border game, if a card has a black border that frames the picture on the card, the child is asked to sort the card according to color rule; but if the card does not a have black border, child is asked to sort according to shape rule. This game is played with 14 cards. Two of those are training cards and they both have blue elephants on them; however, one is bordered and the other is not. Training cards aim to show children that two cards that are identical regarding color and shape but differ on whether they have a border or not would be sorted differently. The remaining 12 cards (3 blue elephant-with border, 3 blue elephant-no border, 3 red car-with border, 3 red car-no border) are presented to children in an order so that more than 2 same cards would not show up consecutively. Children are reminded of following rule before each card: *"Now we are playing the border game. In the border game, if there is black border surrounding the card, you should play color game. If there is no black border on the*

card, you should play shape game. This one has/does not have a border. Where does it go?"

Each correct sorting gets 1 point, sorting all of the cards correctly results in 12 points. A child can be considered as successful in border game if s/he gets at least 9 points out of 12.

4.1.2.3 Language Assessment

In the study, children's language abilities were tested with Türkçe İfade Edici ve Alıcı Dil Testi (TIFALDI) which is a standardized test assessing both receptive and expressive language abilities of 2 to 12-years old children in word level (Berument & Güven, 2013).

4.1.2.3.1 Receptive Language Test

In this task, children are asked to identify the correct picture that depicts the target word. Each target word is presented with 3 foil pictures. There is a total of 104 target words. For example, if the target word is "dog", the child sees 4 pictures: a bird, a dog, a duck, and a cat; and s/he is asked "*Which picture shows a dog?"* or "*Which one of those is a dog?".*

The categories in the task include animals, flowers, kitchen utensils, activities, vegetables and fruits, toys, furniture, jobs, household items, body parts, shapes, objects, and emotions. The words are ordered so that they get harder progressively. The test begins with items that are appropriate for child's age. If the child fails during the first 8 words, earlier items are also asked until s/he gets 8 right answers consecutively so that a base score can be determined for the child's performance. The

test is ended if a child makes 8 mistakes in 10 words; otherwise all target words are asked to children. This subtest yields a raw score that is comprised of number of correct answers and a standardized score that is calculated based on child's age and number of correct answers.

4.1.2.3.2 Expressive Language Test

In this task, there are 80 target words. Children are showed these pictures of these words one by one and they are asked to name them. For example, if the target picture shows a dog, the experimenter points to the picture and asks: "*What is this?"* or "*What does this picture show?".*

The pictures used in the task are from diverse categories including animals, flowers, kitchen utensils, activities, vegetables and fruits, toys, furniture, jobs, household items, body parts, shapes, objects, and emotions. Vocabulary that is needed to name the pictures gets progressively harder. The test begins with items that are appropriate for child's age. If the child fails during the first 8 pictures, earlier pictures are also asked until s/he gets 8 right answers consecutively so that a base score can be determined for the child's performance. Testing ends if a child makes 6 mistakes in 8 pictures; otherwise all target pictures are shown to children. This subtest yields a raw score that is comprised of number of correct answers and a standardized score that is calculated based on child's age and number of correct answers.

4.1.3 Procedure

Children whose parents consented to the study were tested in a quiet room at their preschool (or in their homes if the child was being tested at home) and tested alone.

Children's verbal assent was also sought at the beginning of each testing session. Each child was tested on two consecutive days. Testing sessions in both days lasted approximately 20-25 minutes.

We applied the ToM tasks in two separate but consecutive days because some of the ToM tasks were very similar (i.e., only the mental state verbs and objects differed between them) and we wished to prevent children replying similarly without processing the minor differences between the tasks. We also distributed the control tasks (EF and language tasks) across the two testing days. This was done to prevent fatigue.

The distribution of the tasks was determined before data collection (see Table 4 for the distribution of the tasks). In order to stay as close to the administration of Wellman and Liu's (2004) ToM battery in the literature, the original tasks (i.e., DD, DB-Original, KI, EFB-Original, CFB-Original, HE) were always applied together, as the first tasks in either day. In the other day, pragmatically-modified versions of DB, EFB, and CFB tasks were applied as the first tasks. Thirty children were tested on the original tasks on Day 1 and on pragmatically-modified versions on Day 2; and thirty children were tested vice versa. The days in which the original and pragmatically-modified versions of the tasks were applied were purposefully counterbalanced in order to see (and control, eventually) if a learning effect occurred between two days even though no feedback was provided for children in Day 1. However, Mann-Whitney Tests revealed that neither original nor pragmatically-modified tasks were affected by testing order.

On Day 1, EF followed the first set of ToM tasks (i.e., either the original or the modified versions). Testing session of Day 1 was completed with applying the two modified versions of each of DB and EFB tasks, respectively. The places of those tasks in the study were held stable in order to decrease the number of conditions in which children were divided into. One might wonder if the EF task applied immediately before would boost children's performance on ToM tasks; however, to our knowledge, there is no study showing an immediate boosting effect of DCCS on ToM tasks. Additionally, children's extremely poor performance on EFB-Story version (i.e., only four children passed that task) is supporting that EF task did not provided a quick inspiration for ToM tasks. Children were given stickers at the end of Day 1.

Day 2 testing session started with either the original ToM Battery (Wellman and Liu, 2004) or modified versions of DB, EFB, and CFB tasks. After ToM tasks, Receptive and Expressive Language subtests of TIFALDI were applied respectively. Again, children were given stickers at the end of Day 2; but, additionally, a debriefing form that was explaining the aim and the design of the study more thoroughly was sent to parents via preschool management at the end of the second day.

Table 4. Distribution of the Tasks across the Days

Condition 1			Condition 2	
Day 1	Day 2		Day 1	Day 2
DD DB-Original KA EFB-Original CFB-Original HE	DB-*Guess* EFB-*Falsely Think* CFB-*Falsely Think*	or	DB-*Guess* EFB-*Falsely Think* CFB-*Falsely Think*	DD DB-Original KA EFB-Original CFB-Original HE
DCCS	TIFALDI		DCCS	TIFALDI
DB-Evidence EFB-Story			DB-Evidence EFB-Story	

4.2 Results

The current study aimed to investigate whether modifying the belief-related tasks (DB, EFB, and CFB) of Wellman and Liu's (2004) ToM battery in a way that was consistent with Turkish's pragmatics and semantics would affect the performance of Turkish preschoolers. The results will be presented in 6 sections. First, for each task, the percentage of children who passed that task will be presented by age and in total and children's performances on different tasks will be examined. Second, we will report McNemar's tests that looked at whether children benefited from task modifications by comparing children's performances on pairs of the same kind of task, e.g., DB-Original vs. DB-*Guess*. Third, we will report one-way between-subjects ANOVAs which investigated if there were differences in Total ToM performance by age, as literature has suggested so far. Fourth, the partial correlations (age controlled) between ToM, EF, and language abilities will be presented and will be compared to relations that have been found in the literature so far. Fifth, we will present the hierarchical regression analyses in which we investigated whether age, EF, and language abilities, and their interactions, would predict the change in their ToM abilities. And finally, we will report the Guttman analyses which measure the scalability of ToM tasks used with Turkish children.

4.2.1 Children's Performances on ToM Tasks by Age and In Total

The percentage of children (by age and in total) who passed each ToM task can be seen in Table 5. We want to draw attention to two important points displayed by this table. First of all, contrary to what is expected, children's performances showed consistent improvements as age increased only in 4 tasks (KA, both CFB tasks, and HE); the performance in the other tasks either remained constant (e.g., EFB-Story) or

decreased (e.g., DD, all DB tasks, EFB-Original, and EFB-*Falsely Think*) by age. Secondly, as the total percentages imply, children performed lower in all versions of EFB task compared to both the other FB task (i.e., CFB) and to the HE task which has generally been shown to be the most difficult task with Western samples (e.g., Peterson et al., 2005; Wellman and Liu, 2005). Given that a) the children showed poorer performance than expected on both versions of the EFB tasks, b) EFB tasks were not found to be sensitive to the developmental advances in social cognition for Turkish children, c) in the literature, CFB has been used in place of EFB (albeit because of different reasons; see Wellman and Liu, 2004), and d) CFB and CFB-*Falsely Think* showed the expected developmental pattern, from now on, all total scores will include a version of CFB but no version of EFB. EFB will be analyzed further only in the second section, when comparing task modifications in pairs.

Table 5. Percentage of Children Who Passed ToM Tasks by Age

	3 n=20	4 n=20	5 n=20	Total n=60
DD	65%	90%	75%	76.7%
DB-Original	35%	40%	35%	36.7%
DB-*Guess*	30%	50%	30%	36.7%
DB-Evidence	60%	70%	40%	56.7%
KA	20%	85%	100%	68.3%

Table 5 (cont'd)				
CFB-Original	15%	15%	75%	35%
CFB-*Falsely Think*	5%	30%	65%	33.3%
EFB-Original	10%	5%	15%	10%
EFB-*Falsely Think*	15%	10%	25%	16.7%
EFB-Story	0%	10%	10%	6.7%
HE	0%	15%	50%	21.7%

4.2.2 Did Children Benefit from Task Modifications?

In order to see if children's performance improved from the original version of one type of task to the modified version of the same task, we used separate McNemar's tests. McNemar's test is a type of chi-square that is appropriate to use when the categorical data at hand is collected from matched or repeated samples. We conducted 7 separate McNemar's tests on the entire sample and compared children's performances in two different versions of the same type of task: DB-Original vs. DB-*Guess*, DB-Original vs. DB-Evidence, DB-*Guess* vs. DB-Evidence, CFB-Original vs. CFB-*Falsely Think*, EFB-Original vs. EFB-*Falsely Think*, EFB-Original vs. EFB-Story, and EFB-*Falsely Think* vs. EFB-Story. Among all McNemar's tests, only 2 of them indicated a significant difference between children's performance. Specifically, children performed significantly better in DB-Evidence (56.7%) version compared to both DB-Original (36.7%) and DB-*Guess* (36.7%) version of the DB task, $p = .012$

and $p = .008$, respectively. However, contrary to what is expected, children did not perform differently on the DB-Original vs. the DB-*Guess* versions of the task. Similarly, children's performance was not affected by any modification done for the CFB or the EFB tasks.

At this point, we should note that although children's performances in CFB-*Falsely Think* version shows a more gradual development (considering the distribution of successful performance by age), children's overall performance did not differ between two versions of CFB task. From now on, only the CFB-Original version will be included in Total ToM Score as well as the Guttman (sequencing) analysis in order to keep our data as comparable as possible to other ToM scaling studies conducted with Western and Eastern samples (e.g., Shahaeian, Peterson, Slaughter, Wellman, 2011; Wellman, Fang, Liu, Zhu, and Liu, 2006; Wellman and Liu, 2004). Similarly, since the DB-Evidence version is significantly different from both the DB-Original and DB-*Guess* versions and because DB-Original and DB-*Guess* versions did not differ from each other, here on, all further analysis will be presented in two ways: a) with DB-Original and b) with DB-Evidence. Total scores will also be calculated in two ways: a) "ToM Total with DB-Original" which is the aggregate score of children's performances in DD, KA, CFB-Original, HE, and DB-Original tasks; and, b) "ToM Total with DB-Evidence" which is the sum of the performances in DD, KA, CFB-Original, HE, and DB-Evidence.

4.2.3 Are There Differences in Total ToM Performance by Age?

In order to determine whether age had an effect on the total ToM performance of children, as suggested by literature so far (e.g., Happé, 1995; Wellman and Liu,

2004), we conducted two separate one-way ANOVAs with age as the independent factor and "ToM Total with DB-Original" or "ToM Total with DB-Evidence" as the dependent variables. The analysis showed the expected effect of age on both "ToM Total with DB-Original" (F (2, 57) = 21.98, p <.001, η_p^2 = .44) and "ToM Total with DB-Evidence" (F (2, 57) = 19.19, p <.001, η_p^2 = .40). Post-hoc Bonferroni's tests revealed that when "ToM Total with DB-Original" was the DV, 3-year old children (M = 1.35, SD = 0.93) differed from both 4- (M = 2.45, SD = .94) and 5-year old children (M = 3.35, SD = .99), p = .002 and p < .001, respectively. Additionally, 4 – and 5-year old children differed from each other, p = .01. When the "ToM Total with DB-Evidence" was the DV, though, Bonferroni's tests revealed that 3-year old children (M = 1.60, SD = 0.88) differed from both 4- ((M = 2.75, SD = 0.85) and 5-year old children ((M = 3.40, SD = 1.04), p = .001 and p < .001, respectively; however, the difference between 4- and 5-year old children was marginally significant, p = .09.

4.2.4 Partial Correlations across Children's ToM, EF, and Language Abilities

The literature has suggested consistent and moderate to strong relations among ToM, EF, and language abilities of children, e.g., Hughes, 1998. In order to check whether these relations between children's cognitive variables (language and EF) were related to their ToM scores, we conducted partial correlations and controlled for the effect of age. The partial correlations among the two aggregate scores for ToM ("ToM Total with DB-Original" or "ToM Total with DB-Evidence"), EF, and language abilities of children can be seen in Table 6. In the table, EF corresponds to the total number of cards sorted correctly by children in DCCS (out of 24) and receptive and expressive language respectively corresponds to raw receptive and raw expressive language

scores children received in TIFALDI. The total language, though, corresponds to the sum of raw receptive and raw expressive language scores children received in TIFALDI. The partial correlation analyses revealed that "ToM Total with DB-Original" weakly correlated with children's expressive language abilities (r =.29, p = .03) but neither with receptive language abilities (r =.06, p = .63) nor total language (r =.17, p = .19); its correlation with EF was marginally significant, r =.23, p = .09. However, the analyses showed that "ToM Total with DB-Evidence" moderately correlated with children's expressive language abilities (r =.30, p = .02), total language (r =.25, p = .05), and EF (r =.36, p = .005). The significant correlation between language and ToM Total with DB-Evidence was carried through the expressive language score and receptive language abilities were still not correlated with ToM (r =.18, p = .16); however, the correlation coefficient (r) tripled compared to r found with "ToM Total with DB-Original". These findings are suggesting that "ToM Total with DB-Evidence" is a better aggregate score for explaining the expected relations between ToM and EF or ToM and language abilities of children.

Table 6. Age-Partialed Correlations between Two Different ToM-Total and EF and Language Abilities

	ToM Total with DB-Original	ToM Total with DB-Evidence
EF	.23	.36**
Receptive Language	.06	.18
Expressive Language	.29*	.30*
Total Language	.17	.25*

Note. *p < .05, **p < .01 (All significance tests are two-tailed)

4.2.5 Can Control Variables (i.e., Age, EF, and Language) Concurrently Predict Children's Total ToM Scores?

Two hierarchical regression analyses (separate analyses for "ToM Total with DB-Original" and "ToM Total with DB-Evidence") were conducted in order to determine which control variable (i.e., children's ages, EF and language abilities) is better predicting the change in children's Total ToM Scores (Table 7 & 8). Preliminary analyses ensured that there was no violations of normality, linearity, multicollinearity, and homoscedasticity in any of the hierarchical regression analyses. In the first regression analysis, "ToM Total with DB-Original" was entered as DV. Age in months, sum of raw language scores (expressive and receptive language), and total EF scores (out of 24) were entered at step 1. The first model explained 42% of the variance in ToM Total with DB-Original and 44% of the variance in ToM Total with DB-Evidence. Since age was highly correlated with both EF (r = .57, p < .001) and language abilities (r = .71, p < .001) of children, we wanted to know whether its interaction with these variables would add on to variance explained by the independent age, language, and EF variables. For this purpose, we created two interaction terms (i.e., Age-Language Interaction and Age-EF Interaction) and entered them at step 2. After entering the interaction terms, the total variance explained by the model increased to 47.3% for ToM Total with DB-Original, F (3, 56) = 15.24, p< .001. After controlling for the independent effects of age, language, and EF, the two interaction terms explained an additional 5.3% of the variance in ToM Total with DB-Original, R squared change = .05, F change = (2, 54) = 3.83, p = .03. In the final model, Age-Language Interaction marginally predicted the change in ToM Total with DB-Original (β = .25, p =.06) while Age-EF Interaction negatively predicted it (β = -.36, p =.008).

In the second regression analysis, "ToM Total with DB-Evidence" was entered as DV. Again, age in months, sum of raw language scores (expressive and receptive language), and total EF scores (out of 24) were entered at step 1. The first model explained 44% of the variance in ToM Total with DB-Evidence. Similar to the first regression analysis conducted with "ToM Total with DB-Original", we entered two interaction terms (i.e., Age-Language Interaction and Age-EF Interaction) at step 2 in order to predict "ToM Total with DB-Evidence". The interaction terms increased the total variance explained by the model to 49.4%, F (3, 56) = 16.33, p< .001. After controlling for the independent effects of age, language, and EF, the two interaction terms explained an additional 5.4% of the variance in ToM Total with DB-Evidence, R squared change = .05, F change = (2, 54) = 4.13, p = .02. In the final model for ToM Total with DB-Evidence, Age-Language Interaction did not predict any change (β = .18, p =.16) while Age-EF Interaction negatively predicted it (β = -.36, p =.007).

These results suggest that the interaction of children's age and EF abilities predicted their either ToM performances above and beyond the individual effects of these variables. However, while age and EF (only for "ToM Total with DB-Evidence") variables were positive predictors of Total ToM scores, age by EF interaction was a negative predictor of ToM Total (see Figures 2, 3, and 4 for regression plots).

Table 7. Predicting "ToM Total with DB-Original Performance" by Age, Language, and EF

Variables	B	SE B	β	t	Sig
Step 1					

Table 7 (cont'd)					
Constant	-1.55	.67		-2.31	.025
Age	.05	.02	.46	3.17	.002
Language	.01	.01	.12	.79	.434
EF	.04	.03	.18	1.38	.174
Step 2					
Constant	-1.82	.67		-2.73	.009
Age	.06	.02	.48	3.46	.001
Language	.01	.01	.14	1.00	.322
EF	.04	.03	.21	1.64	.106
Age-Language Interaction	.00	.00	.25	1.94	.058
Age-EF Interaction	-.01	.00	-.36	-2.77	.008

Note. Adjusted R^2 = .42 for Step 1, Adjusted R^2 = .47 for Step 2

Table 8. Predicting Children's "ToM Total with DB-Evidence" Performance by Age, Language, and EF

Variables	B	SE B	β	t	Sig
Step 1					
Constant	-.79	.62		-1.26	.211
Age	.04	.02	.32	2.25	.029
Language	.01	.01	.17	1.13	.263
EF	.06	.03	.30	2.40	.020
Step 2					

Table 8 (cont'd)

Constant	-.92	.62		-1.50	.140
Age	.04	.02	.33	2.43	.018
Language	.01	.01	.18	1.31	.195
EF	.07	.02	.35	2.86	.006
Age-Language Interaction	.00	.00	.18	1.41	.164
Age-EF Interaction	-.01	.00	-.36	-2.81	.007

Note. R^2 = .44 for Step 1, Adjusted R^2 = .49 for Step 2

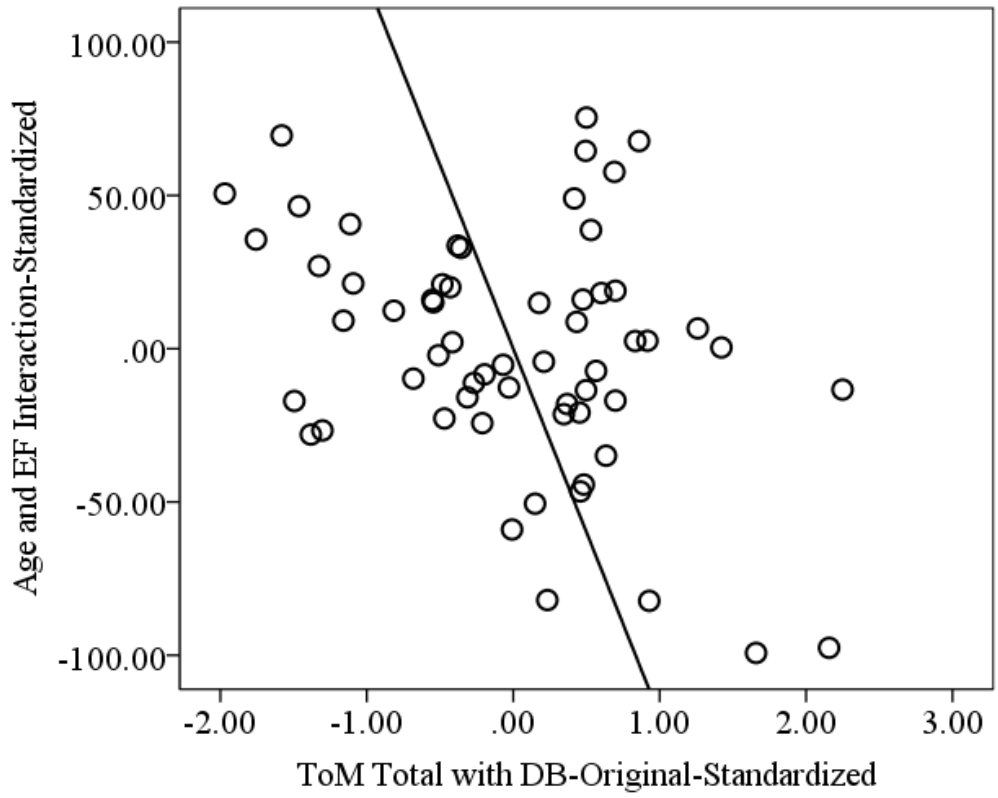


Figure 2. Predicting ToM Total with DB-Original by Age and EF Interaction

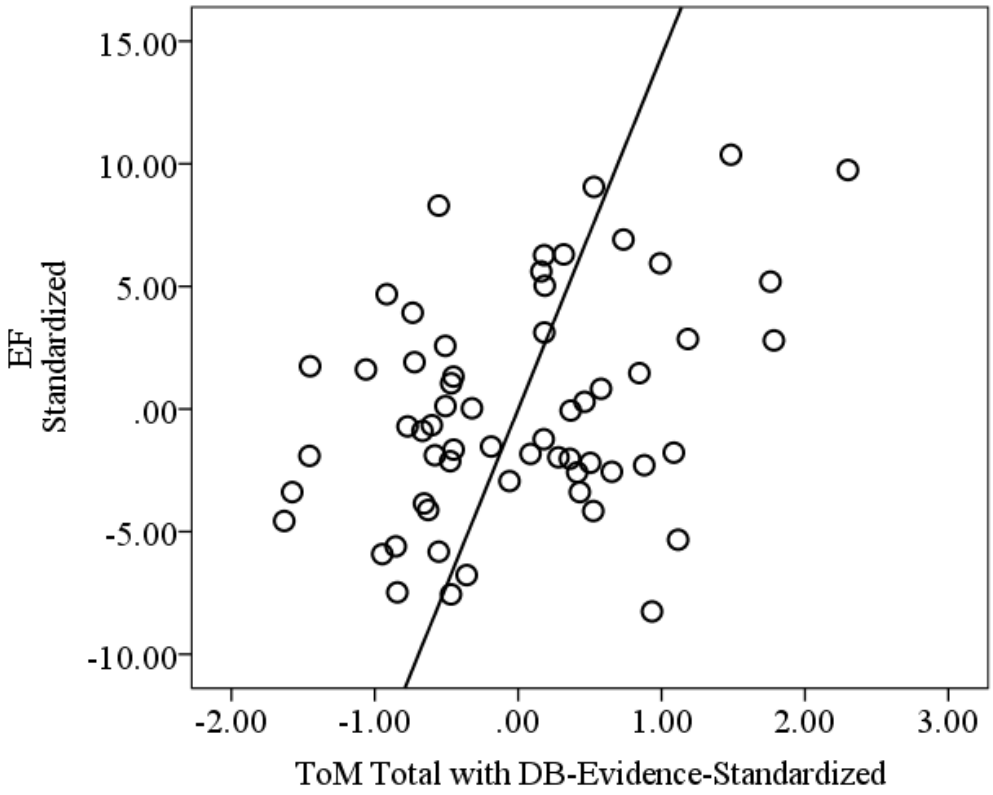


Figure 3. Predicting ToM Total with DB-Evidence by EF

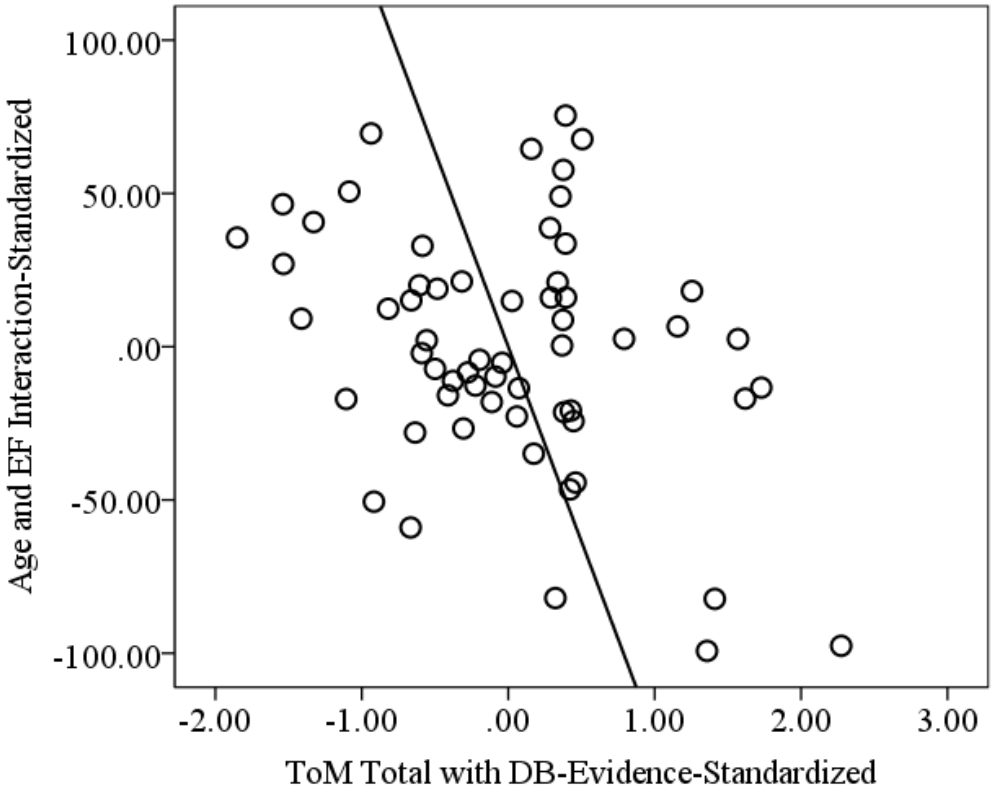


Figure 4: Predicting ToM Total with DB-Evidence by Age and EF Interaction

4.2.6 Sequencing Analysis: Five-item Guttman Scale

One primary aim of this study is to find out whether ToM tasks could be meaningfully sequenced so that they create a pattern from easiest to hardest, based on Turkish children's performances on these tasks. In order to sequence ToM tasks, we used Guttman scaling method described in Green, 1956. This method has been used widely in scaling studies conducted so far, e.g., Etel and Yağmurlu, 2015; Shahaeian, Peterson, Slaughter, Wellman, 2011; Wellman, Fang, Liu, Zhu, and Liu, 2006; Wellman and Liu, 2004. Guttman analysis depends on the logic of perfect Guttman scale in which one positively scored item is always preceded by easier items with positive scores. In this ideal case, if we know one person's score and the pattern formed by the items, we would be able to know how this person answered all items in that scale. Based on the difference between the ideal case and the case at hand, Guttman analysis provides us with "the estimation of reproducibility given knowledge of person scores, that is, the extent to which item responses fit the ideal patterns." (Wellman & Liu, 2004, 532) Two values are produced by a Guttman analyses: a) the *coefficient of reproducibility* corresponds to how well a scale can be estimated from individuals' scores and it is considered as significant if higher than .90; b) the *index of consistency*, which tells us that "whether the observed coefficient of reproducibility was greater than what could be achieved by chance alone" (532); the values higher than .50 are considered significant. Although the index of consistency is necessary to label a bunch of items as "scalable", it is a conservative measure and was considered optional in Green's (1956) article.

As a first step, we conducted two separate Guttman analyses in which we tried to scale the 4 original tasks of Wellman and Liu's (2004) ToM battery (DD, KA, CFB-

Original, and HE) and either DB-Original or DB-Evidence versions of the DB task. Turkish children's performances on the 5 original tasks was forming the pattern that was found with Chinese and Iranian children before (aka Eastern sequence; Shahaeian, Peterson, Slaughter, Wellman, 2011; Wellman, Fang, Liu, Zhu, and Liu, 2006): DD task was easier than KA task, KA task was easier than the DB task, the DB task was easier than CFB task and HE task was the most difficult task. However, the analyses revealed that it was not possible to estimate the scale by looking at individuals' scores and that the items were not scalable, *coefficient of reproducibility* = .89, *index of consistency* = .14. Similar results were observed when the Guttman analysis was conducted with children's scores on 4 original ToM tasks (DD, KA, CFB-Original, and HE) and one modified task (DB-Evidence). Although it seemed like children were following the Eastern sequence mentioned before, neither the *coefficient of reproducibility* nor the *index of consistency* was significant, .88 and .09, respectively.

Since the Guttman analyses that were conducted with the whole sample failed to yield a reproduceable pattern for sequencing ToM tasks and since there was an effect of age by EF interaction on children's ToM performances, we conducted separate Guttman analyses on children who were from different age groups (3, 4, or 5) or who differed in terms of their EF abilities (passed 1 rule/passed 2 rules). This method enabled us to more deeply investigate the effect of the EF by age interaction on ToM scores as it allowed us to see whether the pattern followed by children depended on their age group or EF abilities. Again, separate Guttman analysis were conducted by acknowledging either DB-Original or DB-Evidence as the DB task.

Six (3 age group x 2 DB tasks) separate Guttman analyses were conducted for age group separation. When the DB-Original was used as the DB task, 3-years old children fit the Western pattern; i.e., DD→DB→ KA→ CFB → HE (*coefficient of reproducibility* = .94, *index of consistency* = .21); 4-years-olds fit the Eastern pattern, i.e., DD→KA→ DB→ CFB → HE (*coefficient of reproducibility* = .94, *index of consistency* = .18); and, 5-year-olds followed neither the Western nor the Eastern pattern, they showed a new sequence of tasks in which knowledge access came first, i.e., KA→DD→ DB→ CFB → HE (*coefficient of reproducibility* = .90, *index of consistency* = .00). When DB-Evidence was taken as the DB task, similar results were observed. Accordingly, 3-years-olds fit the Western pattern, *coefficient of reproducibility* = .91, *index of consistency* = -.09; 4-years-old children fit the Eastern pattern, *coefficient of reproducibility* = .93, *index of consistency* = .04; and, 5-year-olds fit the new sequence, *coefficient of reproducibility* = .90, *index of consistency* = .05. We should note that, although all reproducibility coefficients were significant, none of the consistency indexes yielded significant results. It means that although it was possible to estimate the scale by looking at individuals' performances, the items were not scalable from easiest to hardest.

In addition to dividing participants by age, we also divided them according to their success on EF task. This method enabled us to see the most probable ToM pattern in one level of EF. Children who passed only 1 rule (color rule) of DCCS formed one group (n = 33) and children who passed 2 rules (color and shape rules) formed another group (n = 23). We did not include 4 children who passed all 3 rules of DCCS in any group since these children were so few to form an analyzable separate group.

Four (2 EF group x 2 DB tasks) separate Guttman analyses were done for two EF groups (1-rule-passed vs. 2-rules-passed). (see Table 9 for the ToM scales found for separate EF groups) When the DB-Original was taken as the DB task, children who passed 1 rule fit the Eastern pattern, i.e., DD→KA→ DB → CFB → HE (*coefficient of reproducibility* = .92, *index of consistency* = .17); however, children who passed 2 rules in EF task showed a completely new pattern in which DB-Original was the hardest task, even harder than CFB and HE emotion tasks, i.e., DD→KA → CFB → HE → DB (*coefficient of reproducibility* = .91, *index of consistency* = .12). When DB-Evidence was selected as the DB task, though, more consistent results were observed. Specifically, children who passed 1 rule fit the Western pattern, *coefficient of reproducibility* = .90, *index of consistency* = -.03; and the children who passed 2 rules fit the Eastern pattern, *coefficient of reproducibility* = .90, *index of consistency* = -.03). We should note that, although all reproducibility coefficients were significant, none of the consistency indexes yielded significant results. It means that although it was possible to estimate the scale by looking at individuals' performances, the items were not scalable from easiest to hardest.

Table 9. ToM Scales by Different Version of The DB Task and EF Groups

	DB-Original		DB-Evidence	
	Passed 1 Rule	Passed 2 Rules	Passed 1 Rule	Passed 2 Rules
Easiest	DD	DD	DD	DD
	KA	KA	DB	KA
↓	DB	CFB	KA	DB
	CFB	HE	CFB	CFB
Hardest	HE	DB	HE	HE

4.2.7 Summary of Results and Discussion

Study 3 investigated whether Turkish children's performance on pragmatically modified versions of belief-related ToM tasks would differ from their performance on the original versions of these tasks and whether the modified versions constituted better alternatives to measure Turkish children's understandings of diverse and false beliefs. For this purpose, we designed a within-subject study in which all children were tested on each version of the tasks. Preliminary descriptive analyses provided us with some unexpected results: in the DD task, and all versions of the DB and the EFB tasks. Specifically, the percentage of children who passed these tasks was lower in older children as compared to younger children. Among these tasks, EFB tasks showed especially surprising results for three reasons: (1) these tasks failed to show the expected developmental changes, (2) children performed especially poorly regardless of the version, and (3) these tasks were not correlated with the CFB task. Therefore, we decided not to include any version of the EFB task in further analyses. Children's false belief understanding was evaluated in terms of their performance on CFB-Original task for the rest of the study. Among the two versions of CFB tasks, we decided to include the original version (CFB-Original) rather than its modified counterpart for three reasons: a) children were not differentially successful in the modified versus the original version of the task b) in the case of no difference in performance we opted for the original version for the sake of comparability with other studies in the literature; and c) Study 1 had revealed that Turkish adults were more likely to use the "*think"* verb in the target question of the CFB task, hence, CFB task with "*think"* verb could indeed have been the more appropriate choice for Turkish speakers.

For the 3 versions of the DB task the results revealed that Turkish preschoolers benefited differently from these versions. Compared to the DB-Original and the DB-*Guess* versions, Turkish children performed better when diverse beliefs were supported by perceptual evidence (i.e., DB-Evidence) and when these beliefs were expressed with the "*think*" verb. Since children's performance in the DB-Evidence version differed from the other DB versions and since DB-Original and DB-*Guess* versions did not differ at all, we conducted all remaining analyses in two ways: one with DB-Evidence version and the other with DB-Original version. Even the total ToM score was calculated twice with different versions of the DB task. Analyses conducted with two total ToM scores showed that both "ToM Total with DB-Original" and "ToM Total with DB-Evidence" revealed the expected effect of age on children's ToM abilities. However, correlation analyses done to check the expected relations between children's cognitive abilities (i.e., ToM, EF and language) revealed that "ToM Total with DB-Evidence" showed higher correlations with the other cognitive abilities. The regression analyses showed that while age and EF (only for "ToM Total with DB-Evidence") abilities of children predicted their ToM abilities positively, age by EF interaction was a negative predictor of both total ToM scores. These results suggest that the independent variables of age and EF are able to explain significant amount of increase in children's ToM abilities; however, the effect of children's EF on their ToM abilities is dependent on their age.

Interestingly the total language score did not predict the variance in children's total ToM scores that included DB-Evidence. This pattern of results was the same whether we used the standard or raw scores. The participants of Study 3 had higher than average language abilities, given that the means for the standard language scores of

children in our sample was one standard deviation higher than the sample used in the norm study (Berument & Güven, 2013). This was true both for receptive language (*M* = 120.02) and expressive language (*M* = 119.88). The lack of variability in language abilities may explain the absence of expected relations between the language score and the total ToM scores in the regression analysis.

Based on the results of the regression analyses, in addition to the routine scaling analysis on the entire data, we carried out separate Guttman analyses for different age groups (i.e., 3-, 4-, and 5-year-olds) and EF categories (children who passed 1 rule vs. 2 rules in DCCS) in order to better detect the effect of these variables on the scaling of tasks. Guttman analyses on the entire data did not reveal a scalable pattern, however, the Guttman analyses on the different age and EF groups revealed that the most probable developmental progression of ToM tasks varied by age or EF category. On the one hand, the different sequences found for different age groups displayed a consistent pattern and showed that older children found the KA task easier but DD and DB tasks harder. These results were independent of the version of the DB task's version, i.e., whether DB-Original vs. DB-Evidence task was taken as the DB task. On the other hand, the sequences found for separate EF groups differed according to which DB task was put in the analysis. Accordingly, when DB-Original task was taken as the DB task, children who passed 1 rule vs. 2 rules in DCCS differed substantially in terms of the scale: DB-Original was scaled as the third easiest task for children who passed 1 rule whereas, for children who passed 2 rules in DCCS, DB-Original was the hardest task. This pattern is not consistent with any of the patterns that have been found so far in the literature. This pattern also does not support the proposed conceptual differences between ToM tasks since the Theory Theory position

is appreciating that diverse beliefs precedes understanding of false beliefs (Wellman, 1990; Wellman, 2014; Wellman & Liu, 2004). When the DB-Evidence version was taken as the DB task, more consistent and plausible patterns were revealed for both of the EF groups. Specifically, children who passed 1 rule fit the Western pattern in which the DB task was scaled as the second easiest task. The children who passed 2 rules passed KA before DB-Evidence. However, DB-Evidence was easier than the FB task. These results imply that DB-Evidence task is a more accurate measure of Turkish children's ability to appreciate diversity of beliefs.

CHAPTER 5

DISCUSSION

One of the defining features of human social cognition is our ability to attend to and make meaning of mental states that are not readily observable. In order to function effectively in the social world, we regularly have to make inferences about others' (and sometimes retrospectively, our own) mental states based on information available to different parties, their motivations and resultant emotional and cognitive states. Wellman and Liu (2004) have contributed greatly to our understanding of the development of ToM by creating a set of tasks (i.e., the ToM Battery) that track children's progression from simpler to more complex abilities. Wellman takes a Theory Theory perspective in which children are conceptualized as developing ToM concepts (e.g., belief: diverse beliefs, false beliefs) through hearing adults label these concepts in everyday situations. This progression, in Wellman's point of view, is partly universal and partly cultural. He asserts that children will pass tasks that tap diversity (of desires and beliefs), before they appreciate that others' behavior may be governed by false beliefs. This progression, he argues, is universal. He does leave room for cultural variability, in ways that do not change this general sequence.

The current research sought to investigate the cultural variance found between children from English-speaking Western cultures in comparison to those from Easterncultures (e.g., Turkish, Iranian and Chinese children). At this point, the fact

that these two clusters of cultures show different developmental progressions is well-established (e.g., Peterson et al., 2005; Selçuk et al., 2018; Shahaeian et al., 2011; Wellman et al., 2006; Wellman & Liu, 2004). However, there are two general question that remain unanswered about the observed cross-cultural variability.

The first of these questions is about the variability of ToM sequence observed within cultures. Although all children regardless of culture seem to show roughly similar developmental trajectories (diversity precedes false beliefs) in aggregate data (i.e., when 3-, 4-, and 5-year-olds are pooled together), when we look at children's performance by age there seems to be different patterns. Specifically, variability is observed within samples where age group shows one pattern and the rest show another. For example, Selçuk et al. (2018) have shown that age negatively predicted children's adherence to the Western pattern which means that younger Turkish children were more likely to pass the DB task before the KA task whereas older children were more likely to show the reverse pattern. Ilgaz et al. (2019) have shown that when the data is divided into different age groups and analyzed separately, different age groups show different patterns. In line with Selçuk et al.'s (2018) findings younger children are more likely to pass the DB task before the KA and FB tasks whereas older children pass KA tasks earlier than the younger children.

A second unanswered question concerns the success rates in of children in different ToM Battery Tasks. Turkish children's performance in DB, CFB, and EFB are all lower than Western and Chinese children (see Table 2). In the current study, we argued that the reason for this observed difference could be due to the possibility that the direct translations of the tasks were insensitive to the mainstream pragmatic uses

of MSVs in everyday language which may make the tasks more challenging for Turkish-speaking children. Hence, we aimed to use more pragmatically suitable alternatives to the original "*think*" verbs to see whether children's performance would improve.

It is important to state that these modifications were not merely methodological considerations. We take a socio-cultural framework to ToM development where children are conceptualized as constructing their ToM understanding through social interactions. A superficial reading of this perspective may make it seem like it is no different from Theory Theory that also proposes that children attend to ToM words and match them to concepts, eventually creating adult-like understanding of how peoples' behaviors relate to their mental states. The point of difference between these two stances is that the socio-cultural framework argues for the emergence of mental state concepts as embedded within children's experiential contexts (i.e., culture, language). This perspective is more sensitive to variations between cultures since it allows for the consideration that the same verb, in this case "*to think*" may imply slightly different circumstances that warrant its use. We hypothesized that these variances may underlie some of the results in the literature that are difficult to interpret. To this end, we conducted 3 studies, 2 with adults and 1 with preschoolers to investigate whether Turkish speaking adults' intuitions about mental state verbs may be different than the ones traditionally used in the English versions of the tasks, whether their performance is affected by some of these manipulations and finally whether preschoolers would show differential performance based on the pragmatics of the task.

Study 1 showed that adults, in fact, showed different intuitions about which MSVs were more appropriate to be used in the wording of the task vignettes. As we expected, they were more likely to choose the verb "*guess*" when there was no reason to favor one option over another. Adults' preferences for the verb "*think*" vs. "*falsely think*" was not uniform across tasks. Specifically, they preferred the verb "*falsely think*" for the abbreviated and original version of the EFB tasks but preferred the verb "*think*" for CFB. We speculated that the existence or lack of evidence may be the reason for these preferences. However, in the case of the EFB-Narrativized task where there was verbal evidence for the reasons of the protagonist's false belief, Turkish adults did not show a clear preference for either verb. A notable limitation of this study is that whether the participants were parents themselves were not controlled for. In fact, given the recruitment process the study includes many college age adults. It would be pertinent for future studies to exclusively target parents of preschool age children.

In the second study where we investigated whether the task vignette included the verb "*think*" vs. the verb "*falsely think*" in FB tasks (i.e., CFB, EFB-Original, EFB-Narrativized) we saw no difference in the reaction times for correct answers either when the tasks were analyzed separately or when their scores were aggregated according to the verb. However, quite unexpectedly, we saw that adults made mistakes in these tasks. These mistakes did not meaningfully favor a particular version of the task when the tasks were analyzed separately. However, when the scores were aggregated according to the verb, Turkish speaking adults had more correct answers in the "*falsely think*" version of the FB tasks as compared to the "*think*" versions. These results collectively showed that, at least for FB tasks, whether

adults believe the verb "*think*" versus the verb "*falsely think*" is more appropriate to use in FB vignettes, they nevertheless perform better when the characters' false belief is explicitly marked as is custom in the Turkish language. However, for those adults who perform correctly in these tasks verb usage did not result in slower processing. The unexpected results regarding accuracy, made us question whether it would have been more beneficial to include an EF task for these adults to investigate whether EF has any effect on accuracy. The unexpected rate of errors in the data could potentially be explained due to individual differences on key cognitive abilities. Future work with adults, may benefit from including cognitive measures that target abilities associated with ToM.

The main research question of this thesis was whether children's performance would significantly be affected by the usage of mental state verbs (i.e., "*guess*" vs. "*think*" and "*think*" vs. "*falsely think*") as well as whether it would be affected by if the verb "*think*" is used with an evidential basis or not. Study 3 showed that children performed better when the verb "*think*" was used with an evidential basis in the DB task as compared to the other versions. However, the task modifications for CFB and EFB did not result in performance differences. In fact, based on children's exceptionally poor performance in the EFB task we decided not to use it in further analysis. The total ToM scores including DB-Evidence showed stronger correlations with the associated cognitive measures (i.e., EF, Receptive Language and Expressive Language). Regression analysis that investigated predictors of children's total ToM scores with the DB-Evidence showed that age and EF were positive predictors whereas an age by EF interaction was a negative predictor. This meant that the effect of age on children's total ToM scores depended on their EF scores. In order to

investigate the details of this interaction we conducted two scaling analysis with the ToM Battery (i.e., with DB-original scores and with DB-Evidence scores). The results supported our previous decision to include DB-Evidence in place of its original version in the analysis since the expected sequences (i.e., Western sequence, Eastern sequence) was observed when the DB-Evidence was included. In contrast, when DB-original was used, the sequence followed by children who scored higher in EF (i.e., passed 2 rules) did not fit any previous documented pattern, and was difficult to explain.

These results provide partial support for our general argument that the pragmatic features of the tasks may affect children's performances. To elaborate, we argued that in the Turkish language children are gradually socialized to think and talk about beliefs in evidence-based ways. The fact that children benefited from a version of the task that presented evidence for beliefs partially supports this argument. In hindsight, the version that would have been most beneficial for children could be one in which both the verb "*think*" was replaced with the verb "*guess*" and evidence for either option was presented. It is possible that the DB-Evidence performance which in the current study is lower than the English-speaking and Chinese children could be more comparable. Future studies that include such a modification would be complement the current study and enhance our understanding of Turkish speaking children's ToM abilities. A related question regarding the DB-Evidence performance is the lower percentage of 5-year-olds that pass the task in comparison to both of the younger age groups. It is always possible that young children are passing the task for the wrong reasons. In other words, younger children's success may be due to simple associations between the character and the choice (e.g., "Me ➔car, character ➔bushes"). If

children's accurate performance at younger ages is due to false positives, a decrease in performance with age could be explained. This explanation would become stronger if other relevant epistemological factors is also a contributing factor to children's performance. If older children who have better language abilities as compared to younger children are evaluated on tasks that are not sensitive to the pragmatics of their language (i.e., uses the verb "*think*" in guessing contexts) without the appropriate evidential support, they may underperform.

It is interesting to note that the one modification that actually made a difference did not involve a manipulation of MSVs but epistemological circumstances of the MSV (i.e., DB-Evidence). Wellman notes that in the Chinese culture, mothers use fewer MSWs in general as compared to their Western counterparts. In fact, Lu, Su, & Wang (2008) have shown that Chinese 3-year-olds' reference to others in reminiscing is associated with their ToM performance as opposed to their use of mental state language. In addition, the same study (Experiment 2) showed that children improved their ToM scores after an intervention that prompted them to talk about others with no emphasis on MSVs.

The ToM literature places a disproportionate amount of emphasis on mental state words (e.g., Adrián et al., 2007; Peterson & Slaughter, 2003; Racine, Carpendale, & Turnbull., 2007; Ruffman, et al., 2002; Slaughter, Peterson, & Mackintosh, 2007; Symons, Peterson, Slaughter, Roche, & Doyle, 2005). However, m useful in explicitly labeling mental state situations, are not the only tools available in the linguistic repertoire of languages to express mental activity. Turkish language in particular as an agglutinative language that makes extensive use of morpho-syntactic units (e.g.,

suffixes). To give an example, in Turkish, desires can be expressed via suffixes that demarcate the modality for volitional wishes to express desires (i.e., -e, -a), and modality for volitional suggestions to express intention (i.e., -se, -sa). More pertinent to the current discussion, adverbs such as "*bana/sana göre*" (in my/your perspective) or modal adverbials "*ben-ce/sen-ce*" (according to me/you) are everyday linguistic expression that children hear frequently to describe perspectival information. These uses constitute quick and easy linguistic shortcuts to express thoughts and beliefs that speakers mark with perspectival intent (Göksel & Kerslake, 2004). One notable shortcoming of the current study is its almost exclusive focus on MSVs. The fact that the most notable differences were observed in a version that manipulated features of the epistemological context (i.e., presence vs. absence of evidence) highlights the need for future studies to investigate children's ToM understanding in linguistically more diverse and possibly accurate ways.

The observed problem with the versions of the EFB task where overall children showed extremely poor performance may be attributable to both the adherence to the verb "*think*" and "*falsely think*" in these task vignettes and the absence of visual evidence (i.e., absence of pictorial stimuli that showed the evidence). Future studies could systematically incorporate these features to investigate Turkish children's developing FB abilities. We should note that the CFB task at least eliminates one of these problems since the appearance of the box is suggestive of its contents throughout the task. In other words, in the CFB task the evidence for the protagonist's belief is visually present until the child answers the target question. This may explain why this task shows better ToM progression with Turkish.

To conclude, across 3 experiments, the current study investigated the Turkish children's ToM development as assessed by the ToM Battery (Wellman & Liu, 2004). We adopted a socio-cultural perspective and created modified versions of the tasks that targeted the pragmatic use of MSVs. To the best of our knowledge, this is the first study that takes a close look at the socio-cultural appropriateness of ToM subtasks and systematically manipulates them. The results confirmed some of the hypothesis regarding the role of evidence in children's epistemological decisions in response to ToM questions. However, the current study also directs us to take a more comprehensive and flexible perspective on mental state language. The role of mental state words (nouns: belief, thought, idea and verbs: to think, to believe) may as easily be expressed with grammatically more complex yet semantically rich features of languages. Turkish with its grammatical richness may in fact make use of such linguistic structures often both in adult-directed and child-directed talk. Future studies that include these variations in testing the developmental progression of ToM with Turkish speaking-children would not only contribute to our understanding of ToM development in this sample, but also contribute to theoretical discussions that target the ontology of ToM and mechanisms that underlie its developmental change.

REFERENCES

Adrián, J. E., Clemente, R. A., & Villanueva, L. (2007). Mothers' use of cognitive state verbs in picture- book reading and the development of children's understanding of mind: a longitudinal study. *Child Development, 78*(4), 1052-1067.

Aksu-Koç, A. (1988). *The acquisition of aspect and modality: The case of past reference in Turkish*. Cambridge: Cambridge University press.

Aksu-Koç, A. (1998). Changes in the basis for children's assertions: The acquisition of –DIr. Paper presented at the Ninth International Conference on Turkish Linguistics, Oxford.

Aksu-Koç, A., Ögel-Balaban, H. & Alp, I. E. (2009). Evidentials and source knowledge in Turkish. In S. A. Fitneva & T. Matsui (Eds.*), Evidentiality: A window into language and cognitive development, New Directions for Child and Adolescent Development*, 125, 13–28. San Francisco: Jossey-Bass.

Astington, J. W. (2001). The future of theory- of- mind research: Understanding motivational states, the role of language, and real- world consequences. *Child Development, 72*(3), 685-687.

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory-of-mind development. *Developmental Psychology, 35*(5), 1311-1320. https://doi.org/10.1037/0012-1649.35.5.1311

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. Cambridge, MA: MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition, 21*(1), 37-46.

Berument, S. & Güven, A. G. (2013). Turkish Expressive and Receptive Language Test: I. Standardization, reliability and validity study of the Receptive Vocabulary Sub-Scale. *Türk Psikiyatri Dergisi, 24*(3), 192.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development, 15*(4), 481-498.

Dennett, D. C. (1979). *Brainstorms*. Hassocks, Sussex: Harvester.

Devine, R. T., & Hughes, C. (2016). Measuring theory of mind across middle childhood: Reliability and validity of the silent films and strange stories tasks. *Journal of Experimental Child Psychology, 149*, 23-40.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development, 62*(6), 1352–1366. https://doi.org/10.2307/1130811

Etel, E., & Yağmurlu, B. (2015). Social competence, theory of mind, and executive function in institution-reared Turkish children. *International Journal of Behavioral Development, 39*(6), 519-529.

Fabricius, W. V., & Khalil, S. L. (2003). False beliefs or false positives? Limits on children's understanding of mental representation. *Journal of Cognition and Development, 4*(3), 239-262.

Flavell, J. H. (2007). Theory-of-mind development: Retrospect and prospect. In G. W. Ladd (Ed.), *Appraising the human developmental sciences: Essays in honor of Merrill-Palmer Quarterly*. (pp. 38–55). Detroit, MI: Wayne State University Press.

Flavell, J. H., & Miller, P. H. (1998). Social cognition. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Vol. 2: Cognition, perception, and language* (pp, 851-898). New York: Wiley.

Friedman, O., Griffin, R., Brownell, H., & Winner, E. (2003). Problems with the seeing=knowing rule. *Developmental Science, 6*(5), 505-513.

Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory.*Mind & Language, 7*(1- 2), 145-171.

Göksel, A., & Kerslake, C. (2004). *Turkish: A comprehensive grammar*. Routledge.

Happé, F. G. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*(3), 843-855.

Harris, P. L., Donnelly, K., Guz, G. R., & Pitt-Watson, R. (1986). Children's understanding of the distinction between real and apparent emotion. *Child Development, 57*(4), 895-909.

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*(2), 233-253.

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: longitudinal associations with mental-state talk between young friends. *Developmental Psychology, 34*(5), 1026-1037.

Hughes, C., & Ensor, R. (2007). Executive function and theory of mind: Predictive relations from ages 2 to 4. *Developmental Psychology, 43*(6), 1447-1459.

Ilgaz, H., Allen, J.W.P., & Haskaraca, F.N. (2019, June). When cultural variation is the norm: a critical look at the ToM Scale. Paper presented at the 49th annual meeting of the Jean Piaget Society, Portland, OR.

Johnson, C. N. & Maratsos, M. P. (1977). Early comprehension of mental verbs: think and know. *Child Development, 48*, 1743-1747.

Johnson, C. N. & Wellman, H. M. (1980). Children's developing understanding of mental verbs: remember, know, and guess. *Child Development, 51*, 1095-1102.

Kahraman, O. G. (2012). An investigation of the effects of a theory-of-mind based intervention program on 48–60-month-old children's cognitive perspective taking and prosocial behaviors (Unpublished doctoral dissertation). Gazi University, Ankara.

Kristen, S., Thoermer, C., Hofer, T., Aschersleben, G., & Sodian, B. (2006). Validation of the" Theory of Mind" Scale. *Zeitschrift Fur Entwicklungspsychologie Und Padagogische Psychologie, 38*(4), 190-199.

Kuntoro, I. A., Saraswati, L., Peterson, C., & Slaughter, V. (2013). Micro-cultural influences on theory of mind development: A comparative study of middle-class and pemulung children in Jakarta, Indonesia. *International Journal of Behavioral Development, 37*(3), 266–273. https://doi.org/10.1177/0165025413478258

Lalonde, C. E., & Chandler, M. J. (1995). False belief understanding goes to school: On the social-emotional consequences of coming early or late to a first theory of mind. *Cognition & Emotion, 9*(2-3), 167-185.

Lee, K., Olson, D.R., & Torrance, N. (1999). Chinese children's understanding of false beliefs: The role of language. *Journal of Child Language, 26*, 1–21.

Lu, H., Su, Y., & Wang, Q. (2008). Talking about others facilitates theory of mind in Chinese preschoolers. *Developmental Psychology, 44*(6), 1726-1736.

Macnamara, J., Baker, E. & Olson, C. L. (1976). Four-year-olds' understanding of pretend, forget, and know: evidence for propositional operations. *Child Development, 47*,62-70.

Marcovitch, S., O'Brien, M., Calkins, S. D., Leerkes, E. M., Weaver, J. M., & Levine, D. W. (2015). A longitudinal assessment of the relation between executive function and theory of mind at 3, 4, and 5 years. *Cognitive Development, 33*, 40-55.

Meins, E., Fernyhough, C., Arnott, B., Leekam, S. R., & de Rosnay, M. (2013). Mind- mindedness and theory of mind: Mediating roles of language and perspectival symbolic play. *Child Development, 84*(5), 1777-1790.

Meltzoff, A. N. (2002). Imitation as a mechanism of social cognition: Origins of empathy, theory of mind, and the representation of action. In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 6-25). Oxford: Blackwell.

Miller, C. A. (2006). Developmental Relationships between language and theory of mind. *American Journal of Speech-Language Pathology, 15*(2), 142–154. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=EJ902769&site=eds-live

Miller, S. A. (2012). *Theory of mind: Beyond the preschool years*. New York: Psychology Press.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind: Meta-analysis of the relation between language ability and false- belief understanding. *Child Development, 78*(2), 622-646.

Miscione, J. L., Marvin, R. S., O'Brien, R. G. & Greenberg, M. T. (1978). A developmental study of preschool children's understanding of the words 'know' and 'guess'. *Child Development, 42,* 1107-1113.

Moore, C., Bryant, D. & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development, 60,* 167-171.

Moore, C., Pure, K. & Furrow, D. (1990). Children's understanding of the modal expression of speaker certainty and uncertainty and its relation to the development of a representational theory of mind. *Child Development, 61*, 722-730.

Moses, L. J. & Tahiroğlu, D. (2010). Clarifying the relation between executive function and children's theories of mind. In: Carpendale, J., Iarocci, G., Müller, U., Sokol, B., Young, A. (Eds.), *Self-and Social Regulation: Exploring the Relations between Social Interaction, Social Cognition, and the Development of Executive Functions* (pp. 218–231). Oxford University Press, Oxford.

Nelson, K. (1996). *Language in Cognitive Development: The emergence of the mediated mind.* Cambridge, UK: Cambridge University Press.

Olson, S. L., Liu, D., Kerr, D. C., & Wellman, H. M. (2003). Social and behavioral outcomes of children's theory of mind development. Unpublished manuscript.

Özoran, D. (2009). Cognitive development of Turkish children on the relation of evidentiality and Theory of Mind (Unpublished doctoral dissertation). Middle East Technical University, Ankara.

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development, 65*(4), 1228-1238.

Perner, J., & Wimmer, H. (1988). Misinformation and unexpected change: Testing the development of epistemic-state attribution. *Psychological Research, 50*(3), 191–197. https://doi.org/10.1007/BF00310181

Peterson, C., & Slaughter, V. (2003). Opening windows into the mind: Mothers' preferences for mental state explanations and children's theory of mind. *Cognitive Development, 18,* 399–429.

Peterson, C. C., Wellman, H. M., & Liu, D. (2005). Steps in theory- of- mind development for children with deafness or autism. *Child Development, 76*(2), 502-517.

Racine, T. P., Carpendale, J. I., & Turnbull, W. (2007). Parent–child talk and children's understanding of beliefs and emotions. *Cognition and Emotion, 21*(3), 480-494.

Ruffman, T., Slade, L., & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory- of- mind understanding. *Child Development, 73*(3), 734-751.

Schneider, W., Lockl, K., & Fernandez, O. (2005). Interrelationships among theory of mind, executive control, language development, and working memory in young children: A longitudinal analysis. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development: Interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 259–284). Mahwah, NJ: Erlbaum.

Selcuk, B., Brink, K. A., Ekerim, M., & Wellman, H. M. (2018). Sequence of theory-of-mind acquisition in Turkish children from diverse social backgrounds. Infant and *Child Development*. https://doi.org/10.1002/icd.2098

Shahaeian, A., Peterson, C. C., Slaughter, V., & Wellman, H. M. (2011). Culture and the sequence of steps in theory of mind development. *Developmental Psychology*, *47*(5), 1239–1247. https://doi.org/10.1037/a0023899

Slaughter, V., Dennis, M. J., & Pritchard, M. (2002). Theory of mind and peer acceptance in preschool children. *British Journal of Developmental Psychology*, *20*(4), 545-564.

Slaughter, V., Peterson, C. C., & Mackintosh, E. (2007). Mind what mother says: Narrative input and theory of mind in typical children and those on the autism spectrum. *Child Development, 78*(3), 839-858.

Symons, D. K., Peterson, C. C., Slaughter, V., Roche, J., & Doyle, E. (2005). Theory of mind and mental state discourse during book reading and story-telling tasks. *British Journal of Developmental Psychology, 23*, 21-102.

Tamer-Gencer, Z. (2011). Çok kültürlü toplumlarda iletişim: Divriği örneği. *Gümüşhane Üniversitesi İletişim Fakültesi Elektronik Dergisi, 1*(1), 186-205.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origin and role in development* (pp. 103–130). Hillsdale, NJ: Erlbaum.

Vinden, P. G. (1999). Children's understanding of mind and emotion: A multi-culture study. *Cognition & Emotion, 13*(1), 19-48.

Wellman, H. M. (1992). *The child's theory of mind*. The MIT Press.

Wellman, H. M. (2002). Understanding the psychological world: Developing a theory of mind. In U. Goswami (Ed.), *Handbook of childhood cognitive development* (pp. 167-187). Oxford, England: Black.

Wellman, H. M. (2014). *Making minds: How theory of mind develops*. Oxford University Press.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta- analysis of theory- of- mind development: The truth about false belief. *Child Development, 72*(3), 655-684.

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science, 17*(12), 1075–1081. https://doi.org/10.1111/j.1467-9280.2006.01830.x

Wellman, H. M., & Liu, D. (2004). Scaling of Theory-of-Mind Tasks. *Child Development, 75*(2), 523-541. https://doi.org/10.1111/j.1467-8624.2004.00691.x

Youngblade, L. M., & Dunn, J. (1995). Individual differences in young children's pretend play with mother and sibling: Links to relationships and understanding of other people's feelings and beliefs. *Child Development, 66*(5), 1472-1492.

Zelazo, P. D. (2006). The Dimensional Change Card Sort (DCCS): A method of assessing executive function in children. *Nature Protocols, 1*(1), 297-301.

APPENDICES

APPENDIX A: DEMOGRAPHIC FORM USED IN STUDY 1 & 2

1. Adınız ve soyadınız:
2. Yaşınız:
3. Anadiliniz:
4. Cinsiyetiniz: [ ] Kadın [ ] Erkek
5. Eğitim durumunuz:
   [ ] İlköğretim
   [ ] Ortaokul
   [ ] Lise
   [ ] Üniversite
   [ ] Üniversiteye devam ediyorum
   [ ] Yüksek lisans
   [ ] Yüksek lisansa devam ediyorum
   [ ] Doktora
   [ ] Doktoraya devam ediyorum
   [ ] Cevap vermek istemiyorum

APPENDIX B: DIVERSE AND FALSE BELIEF VIGNETTES USED IN STUDY 1

Diverse Belief (DB):

1.

Bu Dilek. Dilek köpeğini bulmak istiyor. Köpeği ya çalılıkların içinde ya da arabanın altında.

Senin _____ ne? Köpek nerede?

düşüncen

tahminin

ENG: This is Dilek. Dilek wants to find her dog. Her dog is either in the bushes or under the car. What is your _____? Where is the dog?

□ thought

□ guess

2.

Bu Tolga.Tolga arabasının anahtarını bulmaya çalışıyor. Anahtar ya çantasında ya da bir gün önce giydiği ceketin cebinde. Senin _____ ne? Anahtar nerede?

düşüncen

tahminin

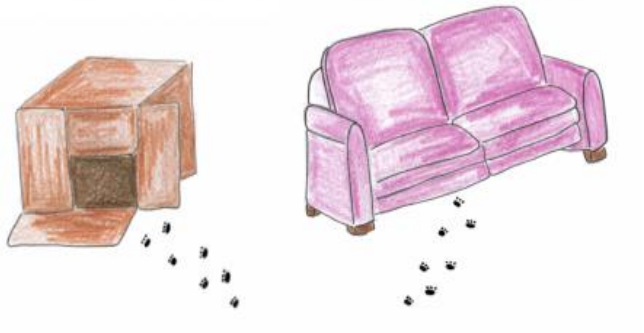
ENG: This is Tolga. Tolga tries to find his car key. His key is either in his bag or in the pocket of the jacket he wore previous day. What is your _____? Where is the key?

□ thought

□ guess

3.

Bu Ceren. Ceren'in kedisi çok kirlenmiş. Ceren kedisini bulup temizlemek istiyor. Pati izlerine göre kedisi ya kutunun içinde ya da koltuğun altında.



Senin_____ ne? Kedi nerede?

tahminin

düşüncen

ENG: This is Ceren. Ceren's cat got really dirty. Ceren wants to find her cat and clean it. Considering the paw marks, her cat is either in the box or under the sofa. What is your _____? Where is the cat?

□ guess

□ thought

4.

Bu Cem. Cem yol tabelalarını takip ederek hiç bilmediği bir şehre gitmeye çalışıyor. Bir yol ayrımında Cem gideceği yerin adının yazdığı iki tane tabela görüyor ama tabelalardan birisi sağı işaret ederken diğeri solu işaret ediyor. Senin _____ ne? Hangi tabela doğru?

tahminin

düşüncen

ENG: This is Cem. Cem tries to drive to a city in which he has never been by following the road signs. In a fork in the road, he sees two signs showing the way to his destination; but, one of them shows the left and the other one shows the right. What is your _____? Which is the right sign?

□ thought

□ guess

Explicit False-Belief (EFB)

5.

Bu Elif. Elif eldivenlerini arıyor. Elif'in eldivenleri ya çekmecede ya da dolapta.

Aslında Elif 'in eldivenleri gerçekten de çekmecede ama Elif eldivenlerinin dolapta olduğunu_____.

düşünüyor

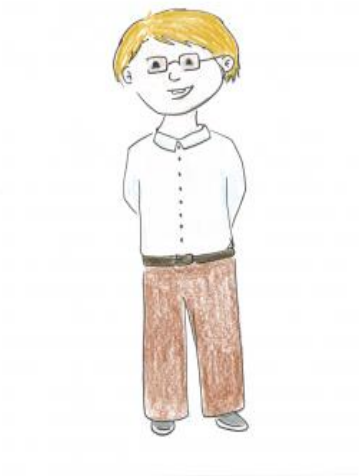
sanıyor

ENG: This is Elif. Elif searches her gloves. Elif's gloves are either in the drawer or in the cupboard. Actually, her gloves are really in the drawer but she _____ that her gloves are in the cupboard.

□ thinks

□ falsely thinks

6.



Bu Ercan. Ercan boya kalemlerini arıyor. Ercan'ın boya kalemleri ya oyuncak kutusunda ya da sırt çantasında.

Aslında Ercan'ın boya kalemleri gerçekten de oyuncak kutusunda ama Ercan boya kalemlerinin sırt çantasında olduğunu _____.

düşünüyor

zannediyor

ENG: This is Ercan. Ercan searches his crayons. His crayons are either in the toy box or in his backpack. Actually, his crayons are really in the toy box but he _____ that his crayons are in his backpack.

□ thinks

□ falsely thinks

Narrativized Explicit False-Belief (Narrativized EFB):

7.

Bu Emre. Emre ders notlarını yazdığı kâğıdı bulmaya çalışıyor. Notlar ya kitabının arasında ya da dosyasının içinde. Aslında notlar gerçekten de kitabın arasında ama Emre notların dosyasında olduğunu _____; çünkü Emre dersten çıkarken notlarını dosyaya koymuştu ama o tuvalete gittiği sırada sınıf arkadaşı notlara bakıp onları kitabın arasına koydu ve bunu Emre'ye söylemeyi unuttu.

zannediyor

düşünüyor

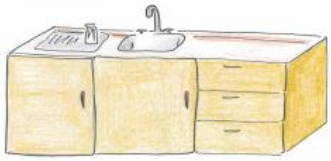
ENG: This is Emre. Emre tries to find his lecture notes. His notes are either in his book or in his folder. Actually, his notes are really in the book but Emre _____that his notes are in his folder because Emre put his notes in the folder after the class. But when he was in bathroom, his friends took his notes from the folder, looked at them, and put them in the book. Then his friend forgot to tell this to Emre.

□ falsely thinks

□ thinks

8.

Bu Fatma. Fatma çikolatasını arıyor. Çikolata ya buzdolabında ya da mutfaktaki çekmecede.



Aslında çikolata kutusu gerçekten de buzdolabında ama Fatma çikolatanın mutfaktaki çekmecede olduğunu _____; çünkü Fatma çikolatalarını en son mutfaktaki çekmeceye koymuştu ama o uyurken annesi çikolatayı erimesin diye çekmeceden alıp buzdolabına koydu ve bunu Fatma'ya söylemeden işe gitti.

sanıyor

düşünüyor

ENG: This is Fatma. Fatma looks for her chocolate. Her chocolate is either in the refrigerator or in the drawer in the kitchen. Actually, her chocolate is really in the refrigerator but she _____ that the chocolate is in the drawer because she put the chocolate in the drawer before she went to sleep. But when she was sleeping, her mother took the chocolate from the drawer and put it in the refrigerator so that it won't melt. Then she went to work without informing Fatma about the situation.

□ falsely thinks

□ thinks

Content False Belief (CFB):

9.

Yukarıdaki kutu bir şeker kutusu ama aslında içinde bir boya kalemi var.

Bu Canan. Canan bu kutunun içinde ne olduğunu hiç görmedi. Peki Canan bu kutunun içinde ne olduğunu_____?

düşünür

zanneder

ENG: This is a candy box but actually it contains a crayon. Canan has never seen what is inside the box. What does Canan _____ that is inside the box?

□ thinks

□ falsely thinks

10.

Bu Can. Can daha önce hiç gitmediği bir eve misafirliğe gidiyor ve orada bir köpek kulübesi görüyor. Ama aslında o kulübede bir tavşan yaşıyor. Can o kulübenin içinde gerçekten ne yaşadığını hiç görmedi. Peki Can kulübenin içinde hangi hayvanın yaşadığını _____?

sanar

düşünür

ENG: This is Can. Can goes to somebody's home in which he has never been before, and he sees a doghouse there. Indeed, in that doghouse, a rabbit lives in. Can has never seen what actually lives in that doghouse. Which animal, Can _____ that lives in that doghouse?

□ falsely thinks

□ thinks

APPENDIX C: FALSE BELIEF TASKS USED IN STUDY 2

Content False-Belief (CFB):

1. Yandaki kutu bir şeker kutusu ama aslında içinde bir oyuncak var. Canan bu kutunun içinde ne olduğunu hiç görmedi. Peki Canan bu kutunun içinde ne olduğunu zanneder? Şeker mi oyuncak mı?

   ENG: This is a candy box but actually it contains a toy. Canan has never seen what is inside the box. What does Canan falsely thinks that is inside the box? Candy or toy?

2. Yandaki kutu bir yara bandı kutusu ama aslında içinde bir boya kalemi var. Can bu kutunun içinde ne olduğunu hiç görmedi. Peki Can bu kutunun içinde ne olduğunu düşünür? Boya kalemi mi yara bandı mı?

   ENG: This is a Band-Aid box but actually it contains a crayon. Can has never seen what is inside the box. What does Can thinks that is inside the box? Crayon or Band-Aid?

Explicit False-Belief (EFB):

3. Bu Elif. Elif eldivenlerini arıyor. Elif'in eldivenleri ya sırt çantasında ya da dolapta. Aslında Elif 'in eldivenleri gerçekten de sırt çantasında. Ama Elif eldivenlerinin dolapta olduğunu zannediyor. Elif eldivenlerini bulmak için nereye bakacak? Dolaba mı sırt çantasına mı?

   ENG: This is Elif. Elif searches her gloves. Elif's gloves are either in her backpack or in the cupboard. Actually, her gloves are really in the backpack, but she falsely thinks that her gloves are in the cupboard. Where will Elif look to find her gloves? The cupboard or her backpack?

4. Bu Ercan. Ercan boya kalemlerini arıyor. Ercan'ın kalemleri ya oyuncak kutusunda ya da çekmecede. Aslında Ercan'ın boya kalemleri gerçekten de

oyuncak kutusunda. Ama Ercan kalemlerinin çekmecede olduğunu düşünüyor. Ercan kalemlerini bulmak için nereye bakacak? Oyuncak kutusuna mı çekmeceye mi?

ENG: This is Ercan. Ercan searches his crayons. His crayons are either in the toy box or in the drawer. Actually, his crayons are really in the toy box but Ercan thinks that his crayons are in the drawer. Where will Ercan look to find his crayons? Toy box or the drawer?

Narrativized Explicit False-Belief (Narrativized EFB):

5. Bu Emre. Emre ders notlarını yazdığı kâğıdı bulmaya çalışıyor. Notlar ya kitabının arasında ya da dosyasının içinde. Aslında notlar gerçekten de kitabın arasında ama Emre notların dosyasında olduğunu sanıyor. Çünkü Emre dersten çıkarken notlarını dosyaya koymuştu ama o tuvalete gittiği sırada sınıf arkadaşı notlara bakıp onları kitabın arasına koydu ve bunu Emre'ye söylemeden sınıftan çıktı. Peki Emre çalışmak için notlarını aradığında nereye bakacak? Dosyasına mı kitabın arasına mı?

   ENG: This is Emre. Emre tries to find his lecture notes. His notes are either in his book or in his folder. Actually, his notes are really in the book but Emre falsely thinks that his notes are in his folder because Emre put his notes in the folder after the class. But when he was in bathroom, his friends took his notes from the folder, looked at them, and put them in the book. Then his friend forgot to tell this to Emre and left class. Where will Emre look when he wants to study his notes? Folder or book?

6. Bu Fatma. Fatma çikolata kutusunu arıyor. Çikolata kutusu ya buzdolabında ya da mutfaktaki çekmecede. Aslında çikolata kutusu gerçekten de buzdolabında ama Fatma çikolatanın mutfaktaki çekmecede olduğunu

düşünüyor. Çünkü Fatma çikolatalarını en son çekmeceye koymuştu ama annesi Fatma uyurken çikolatayı erimesin diye çekmeceden alıp buzdolabına koydu ve bunu Fatma'ya söylemeden işe gitti. Peki Fatma uyanınca çikolata kutusu için nereye bakacak? Buzdolabına mı çekmeceye mi?

ENG: This is Fatma. Fatma looks for her chocolate box. Her chocolate box is either in the refrigerator or in the drawer in the kitchen. Actually, her chocolate is really in the refrigerator, but she thinks that the chocolate is in the drawer because she put the chocolate in the drawer before she went to sleep. But when she was sleeping, her mother took the chocolate from the drawer and put it in the refrigerator so that it won't melt. Then she went to work without informing Fatma about the situation. Where will Fatma, when she wakes up, look to find her chocolate box? Refrigerator or drawer in the kitchen.

APPENDIX D: DEMOGRAPHIC FORM USED IN STUDY 3

1. Adınız ve soyadınız:
2. Çocuğunuzla olan yakınlığınız:
   [ ] Anne
   [ ] Baba
3. Yaşınız:
4. Eğitim durumunuz nedir?

[ ] İlköğretim
[ ] Ortaokul
[ ] Lise
[ ] Üniversite
[ ] Yüksek Lisans
[ ] Doktora
[ ] Diğer: _____

5. İşiniz:

[ ] Tam zamanlı çalışmaktayım
[ ] Yarı zamanlı çalışmaktayım
[ ] Evden yürütmekteyim
[ ] Şu anda çalışmamaktayım
[ ] Okula devam etmekteyim

6. Eşinizin yaşı:
7. Eşinizin eğitim durumu nedir?

[ ] İlköğretim
[ ] Ortaokul
[ ] Lise
[ ] Üniversite
[ ] Yüksek Lisans
[ ] Doktora
[ ] Diğer: _____

8. Eşinizin işi:

[ ] Tam zamanlı çalışmakta
[ ] Yarı zamanlı çalışmakta
[ ] İslerini evden yürütmekte
[ ] Şu anda çalışmamakta
[ ] Okula devam etmekte

9. Çocuk Sayısı:

10. Çocuklarınızın ad(lar)ı ve doğum tarih(ler)i:

_____ _____

_____ _____

_____ _____

_____ _____

11. Evinizin aylık gelir düzeyi:

[ ] 1.650 TL veya daha az
[ ] 1.651 TL - 3.499 TL
[ ] 3.500 TL - 4.999 TL
[ ] 5.000 TL - 6.999 TL
[ ] 7.000 TL veya daha fazla

12. Evinizin aylık gelir düzeyini nasıl değerlendirirsiniz?

[ ] Düşük
[ ] Orta seviyede
[ ] İyi seviyede
[ ] Çok iyi seviyede