THE EFFECT OF LANGUAGE ON GENERIC KNOWLEDGE UNDERSTANDING AND SOURCE RELIABILITY: A DEVELOPMENTAL STUDY

MERVE ATAMAN

BOĞAZİÇİ UNIVERSITY

2018

THE EFFECT OF LANGUAGE ON GENERIC KNOWLEDGE UNDERSTANDING AND SOURCE RELIABILITY: A DEVELOPMENTAL STUDY

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirements for the degree of

Master of Arts

in

Psychology

by

Merve Ataman

Boğaziçi University

2018

The Effect of Language on Generic Knowledge Understanding and Source Reliability: A Developmental Study

The thesis of Merve Ataman has been approved by:

Assist. Prof. Gaye Soley
(Thesis Advisor) _____

Prof. Ayhan Aksu Koç
(Co-Advisor) _____

Prof. Ayşecan Boduroğlu _____

Prof. Aylin Küntay
(External Member) _____

Assist. Prof. Çağla Aydın
(External Member) _____

June 2018

DECLARATION OF ORIGINALITY

I, Merve Ataman, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or a diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature.....

Date.....11.06.2018.....

ABSTRACT

The Effect of Language on Generic Knowledge Understanding and Source Reliability: A Developmental Study

This study investigated the effect of grammaticalized markers of genericity and evidentiality on children's understanding of generalizability of knowledge and source reliability. It explored whether children generalize information conveyed by a statement marked with the Turkish generic marker *-DIr* and the evidential marker *-mIş/-(y)mIş*, and whether this is related to their evaluations of the inferential and the hearsay functions of the evidential form in terms of reliability. For this purpose, a between subjects design was used to test generalizability in the inferential *-mIş*, hearsay *-(y)mIş* and generic *-DIr* conditions, and a within subjects design was used to test source reliability. Children's Theory-of-Mind skills and language competencies were also investigated. A total of 96 monolingual Turkish children, 4-year-olds (N = 48, 20 girls) and 6-year-olds (N = 48, 29 girls) participated in the study. Results showed that 4-year-olds generalized information more when it was conveyed with the generic *-DIr* and did not generalize when it was conveyed with either the inferential or the hearsay uses of *-mIş/-(y)mIş*. However, 6-year-olds generalized the information in all cases. Both 4- and 6-year-olds attributed higher reliability to inferences based on partial observable evidence than to information based on hearsay. Older children attributed higher reliability to inferences than younger children. No significant interaction between source reliability and generalization of information was found. Children who displayed correct understanding of Knowledge Access component of Theory-of-Mind skills were found to attribute higher reliability to inference, but no relation was found with False-Belief understanding.

ÖZET

Bilginin Genellenebilirliği ve Kaynağının Güvenilirliğini Anlamaya Dilin Etkisi: Gelişimsel Bir Çalışma

Bu çalışmada, bilginin genellenebilirliğine ve kaynağının güvenilirliğine işaret eden dil yapılarının, çocukların genellenebilir bilgi ve kaynak güvenilirliği anlayışı üzerindeki etkisi incelenmiştir. Çocukların, Türkçe'deki genelleyici *-DIr* eki ve kaynak belirten *-mIş/-(y)mIş* eki ile edindikleri bilgiyi genelleyip genellemediklerini; ve bu eğilimin, *-mIş/-(y)mIş* ekinin çıkarımsal kullanımı ve duyumsal kullanımı yoluyla elde edilen bilgiye atfettikleri güvenilirlik ile ilişkili olup olmadığı sorgulanmıştır. Bu eklerin bilginin genellenebilirliğine olan etkisi, denekler arası bir dizayn ile çıkarımsal *-mIş*, duyumsal *-(y)mIş*, ve genelleyici *-DIr* koşullarında ölçülmüştür. Kaynağa atfedilen güveni ölçmek içinse denek içi dizayn kullanılmıştır. Çocukların Zihin Kuramı becerileri ve dil yeterlilikleri ise ayrıca değerlendirilmiştir. Toplam 96 Türkçe tekdilli, 4-yaş (N = 48, 20 kız) ve 6-yaş (N = 48, 29 kız) çocuğu çalışmaya katılmıştır. Sonuçlar, 4-yaş çocuklarının *-DIr* eki ile edindikleri bilgiyi genellerken, kaynak belirten *-mIş/-(y)mIş* ekinin çıkarımsal ve duyumsal kullanımlarında bilgiyi genellemeye gitmediklerini göstermiştir. Ancak, 6-yaş-çocukları tüm koşullarda bilgiyi genelleme eğilimi göstermişlerdir. Hem 4- hem 6-yaş-çocukları bir olayın sonuçlarından çıkarım yoluyla edinilen bilgiye bir başkasından duyum yoluyla edinilen bilgiye göre daha yüksek güvenilirlik atfetmişlerdir. Ayrıca, 6-yaş-çocukları çıkarımsal kaynağa 4-yaş-çocuklarının atfettiğinden daha yüksek güvenilirlik atfetmiştir. Bilgi edinilen kaynağa atfedilen güvenilirlik ile bilgiyi genelleme arasında anlamlı bir ilişki bulunamamıştır. Zihin Kuramı yeteneklerinden Bilgi Edinimi bileşenine ilişkin doğru kavrayış sergileyen

çocuklar çıkarımsal kaynağa daha çok güvenilirlik atfederken, Yanlış İnanç bileşeni ile bir ilişki bulunamamıştır.

ACKNOWLEDGEMENTS

First, I would like to express my gratitute to my thesis co-advisor Prof. Ayhan Aksu Koç. Her courses that I took during my undergraduate and graduate years in psychology department shaped my interest in psycholinguistics. I admired her intellectual capacity and endless energy. To be a part of her academic studies was my dream and I achieved it thanks to her support. Her precious advices, theoretical background and constructive criticism helped me to develop myself. I feel so lucky to have such an advisor.

I also owe special gratitute to my thesis advisor Assist. Prof. Gaye Soley for spending her time with me whenever I needed. I could not finish this thesis without her insights, invaluable ideas and important criticisms. Her precious contributions made my thesis possible. Also, I would like to express my special thanks to Prof. Ayşecan Boduroğlu, who kindly agreed to be in the thesis committee, for offering worthful insights on statistics. I am also grateful to Prof. Aylin Küntay who I was lucky work with at Koç University and agreed to be in my thesis committee. I also wish thank to Assist. Prof. Çağla Aydın for being interested in my study. Her positive energy and insights about the operationalizations of the experiments were so precious for me.

I owe special thanks to Treysi Lusi Terziyan and Stefani Terziyan for helping me through the animations that we prepared for the experiments. Without their support, I could not prepare such good animations. I also wish to thank Leman İşbitiren for being a second coder in my thesis.

I also have a dept of gratitude to the children who participated in this study, their parents and teachers.

I would like to thank to all my friends for making this process easier for me and cheering me up in the hardest time. I also wish to thank the members of Koç University Language and Communication Development Laboratory. I was really happy to spend time at this laboratory with its lovely members.

I also owe special thanks to Salih Emre Devrim for being my best friend, listening to my thesis at least thousand times, embracing this thesis as much as me and supporting me all the time. Without him, it would be much more difficult for me.

Last but not least, I am grateful to my mother Nebahat Ataman and my father Engin Ataman for their unconditional love, endless support, and believing in me in all circumstances. Their patience, love and support are some part of my thesis.

To my mother and father,

For their endless love

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDIX FIGURES

CHAPTER 1

INTRODUCTION

The goal of the present study is to explore the effect of language on children's understanding of generic vs. specific knowledge and source reliability. More specifically, it aims to investigate when children are able to differentiate between knowledge that is specific to a particular situation / object and knowledge that is generalizable to other instances of a kind on the basis of information conveyed by language, and whether this understanding is related to their assessment of the reliability of the messages they receive.

Children acquire knowledge through their everyday experiences with people and objects. While through such experience they acquire knowledge about the properties and functions of particular instances of the objects of interaction, human communication offers a unique means to indicate whether the information is generalizable or should remain as a specific knowledge (Prasada, 2000). In other words, linguistic communication and communicative acts enable generic knowledge transmission between individuals (Csibra & Gergely, 2009), thus going beyond direct experience. This means that language is a powerful instrument for learning about what constitutes generic knowledge and what is specific to particular objects and situations.

In addition to the effect of language on the understanding of generic and specific knowledge, communicative acts in pedagogical contexts have also been found to be particularly effective for learning about specific or generalizable information. Butler and Markman (2014) demonstrated that English-speaking children as young as 4 years, generalize the properties of novel objects that are

presented both linguistically and demonstrated ostensively for their benefit in pedagogical contexts more than when they are presented in contexts of an accidental or an intentional activity. Tamm, Çağlar, Aksu-Koç and Csibra (2014) explored whether children would generalize in the absence of an ostensive demonstration of the property if the linguistic cue for generalizability is grammaticalized and, therefore, obligatory in the language as it is in Turkish. They showed that both 4- and 6-year-old Turkish-speaking children do generalize when the property is communicated only linguistically by use of the grammaticalized generic marker -DIr in the absence of an ostensive demonstration. A further result of this study was that Turkish children did not generalize when the linguistic cue used to present the object property was the grammaticalized evidential marker *-mIş/-(y)mIş,* which may express either inferred or hearsay information, depending on context. Therefore, Tamm et al.'s findings raise the question of why generalization is blocked when the evidential form *-mIş/-(y)mIş* is used. Is it because children interpret the property as inferred from partial direct evidence, thus less reliable and therefore not generalizable, or is it because they interpret the property as based on hearsay, which is not direct but secondhand information and therefore not reliable and generalizable? A third possibility is that the inferential use of *-mIş/-(y)mIş* conveys information specific to the particular object and is therefore not generalizable? A further question is whether information taken to be reliable is more likely to be generalized as compared to information taken to be less reliable.

Generalizability of information might be closely related to the reliability of its source regardless of whether it is expressed in generic language or not. As noted above, children do a lot of their learning through interaction with adults and the reliability of information acquired from them depends on their credibility. Research

on children's ability to assess the credibility of informants has shown that children as young as 14 months differentiate between reliable versus unreliable information on the basis of the credibility of the informant (Zmyj, Buttelmann, Carpenter & Daum, 2010). Beyond this, it has been suggested that the language the information is expressed in, may affect the selective trust of children as young as 4 years (Lucas, Lewis, Pala, Wong & Berridge, 2013). Lucas et al. (2010) found Turkish-speaking children – who have evidential markers in their native language – to be advantaged both in selective trust and false belief performance compared to Chinese-speaking children whose language does not have grammaticalized indicators of the mode of knowledge acquisition. They suggest that the presence of linguistic cues such as evidential markers signaling direct vs. indirect access to knowledge inform about the source of information and thus the reliability of that knowledge. The reasoning is, children exposed to an evidential language get experience in mind reading and making inferences about the mental states of others, a skill which reflects itself in high performance on false belief tasks.

The organization of the chapter is as follows. In section 1.1, a brief description of the Turkish evidential *-mIş/-(y)mIş* and the generic *-DIr* suffixes will be presented. In section 1.2, the research on children's understanding of the linguistic markers of information source will be discussed. Section 1.3 will include an overview of the studies on children's assessment of reliability of information. Section 1.4 will take up the developmental relationship between language and Theory of Mind (ToM). Lastly, studies on reliability and generalizability will be examined in section 1.5.

1.1 Description of the Turkish evidential markers *-mIş/-(y)mIş* and the generic marker *-DIr*

Effectiveness of human communication rests on the reliability of the messages conveyed. All languages have devices that allow their speakers to evaluate the reliability of their assertions and all languages allow their speakers to indicate the type of evidence or source on which the knowledge they are asserting is based (Palmer, 2001: 1). However, languages differ in the extent to which these notions are grammaticalized and therefore obligatorily marked. The marking of source of knowledge, or 'evidentiality', for example, is grammaticalized in about one fourth of all languages (Aikhenvald, 2004: 1).

Turkish is a morphologically rich language where notions related to causality, temporality, and modality are marked on the verb.[1] Both evidential notions related to source of knowledge and epistemic notions related to the speaker's evaluation of the reliability of that knowledge are expressed by suffixes attached to the verb (Aksu-Koç, 2016; Aksu-Koç, Ögel-Balaban & Alp, 2009; Aksu-Koç & Slobin, 1986). In Turkish, these notions are treated under the category of 'modality', which involves the expression of the attitude of the speaker with respect to what s/he is asserting. The Turkish modal system can be characterized in terms of Palmer's (2001) framework where evidentiality and epistemic modality are distinct but closely related categories (Aksu-Koç, 2016). Epistemically modalized utterances express the speaker's judgement with respect to the reliability of an assertion, whereas evidential utterances indicate the source for that evaluation (Palmer, 2001: 1).

[1] The affixes that may appear on the verb mark voice, negation, modality, aspect, tense, person and number. All affixes are subject to vowel harmony; they harmonize with the last vowel of the verb stem (Aksu-Koç and Slobin, 1985: 840). The *-mIş/-(y)mIş* suffix may have the forms -mış/-miş/-muş/-müş, the *-DI/-(y)DI* suffix the forms *-dı/-di/-du/-dü* and *-DIr,* the forms *-Dır/-Dir/-Dur/ -Dür.* These alternations are represented by the *'I'*. The *'D'* stands for the *d* vs. *t* consonant alternation.

Turkish is an evidential language because the distinction between direct versus indirect source of knowledge is grammaticalized. There is an obligatory choice between two past tense suffixes *-DI/-(y)DI* and *-mIş/-(y)mIş* (Aksu-Koç & Slobin, 1986). The modally neutral *-DI/-(y)DI* indicates information either obtained through direct experience or accepted as factual, whereas *-mIş/-(y)mIş* indicates information obtained indirectly either through hearsay or through inference from partial observable evidence. For instance, *Küçük köpeğimi çok severmişim,* '(It is said that) I liked my little dog a lot' is a statement based on hearsay, whereas *Bu bardak kırılmış,* '(I infer that) this glass is broken' said after observing a crack in the glass, is a statement expressing inference.

Turkish expresses epistemic modality through a number of verbal affixes such as the abilitative *-Abil* and the aorist *-Ir,* as in *Dikkat et, bardak kırıl-abil-ir,* 'Be careful, the glass may break', thereby expressing possibility or probability. Another affix is *-DIr* which implies speaker certainty when it expresses generic knowledge as in *Kedi vahşidir* 'Cats are wild' and speaker uncertainty when it expresses probabilistic knowledge associated with lower degrees of speaker certainty as in *Ali bizi bekliyordur, acele edelim,* 'Ali must be waiting for us, lets hurry' (examples from Aksu-Koç, 2016 and Sansa-Tura, 1986, p. 145; see also Aksu-Koç & Alıcı, 2000; Aydın & Aksu-Koç, 2015).

Sansa Tura (1986, p. 146) claimed that the Turkish epistemic marker *-DIr* serves functions ranging from certain to non-certain on a continuum. As a certainty marker *-DIr* conveys the meaning of genericity/ factivity, and this use of *-DIr* is found in scientific statements, descriptions, universal truths and generic facts. However, the non-certain uses of *-DIr* conveys the meanings of probability or possibility. In such uses the *-DIr* suffix expresses predictions that depend on the

previous experiences and general knowledge of the speaker. As the examples from Sansa Tura show (1986, p.146), if the speaker says *Bugün cumartesi, televizyonda güzel şeyler vardır,* 'Today is Saturday, there will be good shows on TV', s/he is predicting that there will be good programs on TV based on her/his previous experience. That is, when a speaker uses *-DIr*, s/he does not have access to any observable evidence. Expressed without *-DIr*, the sentence implies certain knowledge on the part of the speaker. Moreover, Sansa Tura (1986) indicated that *muhakkak,* 'certainly' or *belki*, 'maybe' can be used to clarify the strength of non-certain *-DIr* as highly probable or speculative (p. 149). In short, *-DIr* is used to express meanings ranging from facts held to be certain to predictions of different degrees of likelihood.

The verbal affixes of Turkish that are the focus of the present study are the evidential *-mIş/-(y)mIş*[2] with the hearsay and inferential functions and the epistemic *-DIr* with the generic function. As illustrated above, depending on context, the *-mIş/-(y)mIş* form indicates that the basis for the speaker's proposition is either an inference from some partial observable evidence such as a resultant state or is hearsay based on someone else's report (Aksu-Koç, 1988). Again, depending on context, *-DIr* may express the speaker's evaluation of the statement he asserts as certain, or as uncertain and therefore probable. However, in this study, we will investigate only the generic factual function of *-DIr* suffix, and the inferential and hearsay functions of the *-mIş/-(y)mIş* suffix.

[2] The inflection *-mIş* is appended to verbal predicates and may express perfect aspect, past tense and evidential modality depending on the linguistic and/or situational context. The postclitic *-(y)mIş* may be appended to nominal predicates and to already inflected verbs and has an evidential function, either inferential or hearsay, depending on context (Aksu-Koç, 2000). In the present study the affix that figures in the Generalizability task is the postclitic *-(y)mIş* and the affix that figures in the inference statements of the Evidentiary Reliability task is the suffix *-mIş* while in the Hearsay statements of the same task it is the postclitic *-(y)mIş.* Since the present study concerns the functional rather than the formal aspects of the evidential marker, it is always referred to as *-mIş/-(y)mIş.*

1.2 Studies on evidentials as markers of information source

Longitudinal studies of spontaneous speech exploring the acquisition of Turkish have shown that Turkish children acquire evidentiality markers between 1;6 years and 3;0 years (Aksu-Koç, 1988, 1998; Uzundağ, Taşcı, Aksu-Koç & Küntay, 2016), producing them context appropriately. The modally neutral direct experience marker *-DI/-(y)DI* emerges early, around 1;6 years whereas the indirect experience marker *-mIş/-(y)mIş* expressing evidential modality emerges some months later, around 2;0 years. On the other hand, the production of *-DIr* is observed in the second half of the third year (Aksu-Koç et al., 2009).

Experimental studies have shown that successful performance on tasks assessing the production and comprehension of the three forms is highest for *-DI/-(y)DI,* then for *-mIş* (Aksu-Koç, 1988; Ögel, 2007) and *-DIr* (Aksu-Koç & Alıcı, 2000). As Aksu-Koç (1988) indicated, children acquire *-DI/-(y)DI* as the first form of past tense, and direct experience precedes indirect experience. In experiments, 3-year-old children are capable of producing *-DI/-(y)DI* to express direct experience, whereas only half of them use the *-mIş* form for inferences from resultant states. The correct production of inferential *-mIş* is observed for almost all children by 5-years. Ögel (2007) also found that 3- to 6-year-old children use *-DI/-(y)DI* correctly when they report their direct experiences. Four-year-olds produced *-mIş* to describe their inferences from observable results of events whereas only a few 3-year-olds did so. Children's performance on tasks assessing the hearsay use of *-mIş* also increased with age. Five and 6 years old children reported a story they heard in *-DI/-(y)DI* form to another person by using the *-mIş* form, showing evidence for the use of the hearsay function significantly more than 3- and 4-year-old children did. This developmental pattern is similar to the findings of Aksu-Koç (1988). Furthermore,

Aksu-Koç and Alıcı (2000) found that children are significantly more successful in identifying *-DI/-(y)DI* statements as referring to situations directly experienced by the speaker as compared to identifying *-DIr* statements as expressing inferences deduced from their general experience.

Previous experimental research claimed that Turkish speaking children do not produce evidentials above chance level before four years of age, and children aged 6 years are still not like adults in their production and comprehension of evidential markers (Öztürk & Papafragou, 2007). Therefore, Öztürk and Papafragou (2016) conducted a study with 5- and 7-year-olds to see the production and comprehension of evidentials in Turkish. Results showed that production precedes comprehension, and only the oldest group of children could comprehend the functions of the indirect evidential marker *-mIş*. There is an asymmetry between production and the comprehension of the evidentials (Aksu-Koç, 2009; Ünal & Papafragou, 2016). While confirming the findings of longitudinal research for order of emergence, these studies have revealed older ages for successful performance, possibly due to task demands that require explicit, declarative responses (Aksu-Koç, 2009) or due to higher cognitive demands of the comprehension than the production process (Ünal & Papafragou, 2016).

1.3 Studies on reliability

The results of the studies reviewed above have demonstrated that children differentiate between different types of information source at an early age. This ability is crucial for their understanding of the reliability of information. Previous research shows that even 14-month-old infants can differentiate reliable and unreliable sources based on cues such as past accuracy of a person's emotional

signals, confidence expressed by certainty of a person's facial and linguistic expressions, and group membership of the source indicated by the use of the same language as the participant's language (Buttelmann, Zmyj, Daum & Carpenter, 2013; Chow, Poulin-Dubois & Lewis, 2008; Poulin-Dubois, Brooker & Polonia, 2011; Zmyj et al., 2010). Other research with older children (3- and 4-year-olds) was conducted by using accurate and inaccurate labelers to assess source reliability (Koenig, Clement & Harris, 2004; Lucas et al., 2013). In these studies, linguistic ability was necessary to evaluate source reliability since children had to evaluate the informants' reliability based on their past accuracy in labeling familiar objects (Blanco, 2013; Koenig et al., 2004; Lucas et al., 2013). These studies are discussed in detail below.

In a study conducted by Zmyj and colleagues (2010), 14-month-old infants saw an adult performing an instrumental action on a familiar object in a competent or an incompetent way. In a subsequent imitation task, infants watched the same adult turn on a light by putting his forehead on a lamp three times in a hand-free condition (i.e., despite the fact that his hands were free to do so) as in Meltzoff, 1988) and they were asked to play in that set-up. The results indicated that infants imitated this novel action significantly more, if the actor had previously acted competently. Buttelmann and colleagues (2013) further demonstrated that 14-month-olds imitate adults who spoke the native language of the infants (German), but not adults who spoke a foreign language (Russian). The results were interpreted as infants perceiving in-group members as more reliable than out-group members. Furthermore, Chow et al. (2008) demonstrated that 14-month-old infants prefer following the eye gaze of a reliable informant rather than an unreliable one. Infants watched an experimenter expressing happiness while she was looking inside of a container that either had a toy

or was empty. Happy expression for the container with the toy demonstrated reliable looker condition and for the empty container, it demonstrated the unreliable looker condition. Then, the experimenter shifted eye gaze toward a target object located either in front of or behind a barrier. Both infants in the reliable and the unreliable looker conditions followed the experimenter's gaze to the target object located in front of the barrier equally often, but the infants in the reliable looker condition followed the experimenter's eye gaze to the target object located behind the barrier more often than the infants in unreliable looker condition.

Koenig and her colleagues (2004) examined whether 3- and 4-year-old children can differentiate reliable and unreliable informants by using a linguistic task. They had two different informants and a few familiar objects. While one of the informants consistently referred to the familiar objects with their correct names, the other one consistently used incorrect names. Then, children were shown three novel objects that the two informants named differently and they were asked for the correct name of these objects. Both 3- and 4-year-old children endorsed labels offered by the reliable source.

Selective trust has also been claimed to be influenced by the structure of language itself. Lucas and his colleagues (2013) tested 3- and 4 year-old Turkish-speaking children (exposed to grammaticalized evidentiality distinctions in their native language), Chinese-speaking children (with higher executive skills) and English-speaking children (as a control group) in a selective trust paradigm. Children were shown two puppets, "Doggy" and "Ducky", and some familiar toys and foods Ducky consistently named toys correctly and foods incorrectly, while Doggy did the opposite. After this phase, children were presented with three novel objects that were named differently by Doggy and Ducky and they were asked to endorse one of the

labels. The results indicated that compared to Chinese and English children, Turkish children were better able to differentiate the reliability of the informants in terms of their expertise. The authors proposed that this effect might be due to the fact that Turkish children are exposed to a language with evidential markers that obligate the users to attend to the source of knowledge and thereby sensitize them to its reliability.

A number of studies have shown that children are also sensitive to evidentiality markers when endorsing information and that they judge direct experience to be more reliable than indirect experience (Aydın & Ceci, 2009; Fitneva 2008; Fitneva 2009; Matsui, Yamamoto & McCagg, 2006; Papafragou, Li, Choi & Han, 2007). Papafragou, Li, Choi and Han (2007) found that 4-year-old Korean children take information reported with the direct evidence marker rather than the hearsay marker as more reliable. Fitneva (2008) presented 6- and 9-year-old Bulgarian-speaking children with two contradictory stories narrated by two adults who used direct perception versus hearsay expressions, inferential versus hearsay expressions, or direct perception versus inferential expressions. Children were asked whom they believed. Both 6- and 9-year-olds chose the adult who reported firsthand information using direct perception markers more than the adult who reported indirect information using hearsay and inferential markers.

Aydın and Ceci's (2009) findings provide further support for children's sensitivity to linguistic markers of evidentiality when deciding on the reliability of the information. They examined children's proneness to misinformation to demonstrate the relationship between suggestibility and reliability of the information source. Turkish 3- to 6-year-olds, listened to a story on a video. While half of the participants listened to the story told by an adult using the direct evidence marker -

DI/-(y)DI, the other half listened to the same story told by the same adult using the hearsay/reportative marker *-mIş*. After the children listened to music for 10 minutes, they were shown another video by a different adult who narrated the same story with the evidential markers switched. The experimenter showed the child either the original version of the video or the version with some misleading details. Then, the experimenter asked them to choose between an image having misleading detail and an image showing the original detail. If the story had first been presented with direct witness *-DI/-(y)DI* marker, children resisted the misinformation presented with the *–mIş* form compared to misinformation presented with the *-DI/-(y)DI* form. However, if the story had first been presented with indirect *-mIş* marker, children were equally suggestible to directly witnessed or hearsay misinformation. Moreover, older children (mean age = 5.5) tended to take the suggestion of the directly witnessed speaker, regardless of their earlier perspectives. Thus, children are sensitive to linguistic cues indicating different degrees of reliability, and attribute more reliability to direct experience of the source with age.

Matsui, Yamamoto and McCagg (2006) investigated how 3- and 6-year-old Japanese-speaking children comprehend certainty and evidentiality expressed by specific particles versus verbs. Results showed that the ability to differentiate between forms expressing certainty develops earlier than the ability to differentiate between forms expressing types of evidentiality. Furthermore, understanding the implications of particles was earlier than understanding the implications of verbs, and older children did better than younger ones. All children performed better on the contrasts of certainty than on the contrasts of evidentiality. More importantly, 6-year-old Japanese-speaking children also thought that the speaker is more reliable when s/he uses a direct evidential marker compared to an indirect evidential marker.

Children's understanding of expressions of certainty-uncertainty in Turkish was explored by Aksu-Koç and Alıcı (2000), who found that 3- to 6-year-old children were more successful in associating speaker certainty with the use of the direct experience marker *-DI/-(y)DI* as compared to associating non-certainty with the use of *-DIr* marker expressing deductions from everyday knowledge. In another study, when 4- and 6-year-olds had to make a choice between two locations to find a hidden object, both groups preferred statements with forms expressing higher degree of certainty / reliability (*-DI/-(y)DI* and zero marking) to those marked with *-DIr*, with older children being significantly more successful (Aydın & Aksu-Koç, 2015).

In summary, these studies show that children can differentiate between reliable and non-reliable sources through linguistic markers early on. Evidential markers guide children for what to believe or not to believe because they indicate how the knowledge is acquired and children know that direct evidence is more reliable than indirect evidence.

1.4 Theory of Mind (ToM)

Research on children's understanding of types of information source has also focused on whether these abilites are related to the development of Theory of Mind (Aksu-Koç, 2009; Aksu-Koç & Alıcı, 2000; Aksu-Koç et al., 2005; Matsui et al., 2006). Theory of Mind (ToM) refers to the ability to understand the mental states (knowledge and beliefs) of others that may be different from one's own, a capacity crucial for predicting their actions (Carlson, Koenig & Harms, 2013; Wellman & Estes, 1986). A commonly used measure of this capacity is the understanding of false beliefs that require understanding the relation between another person's access to information and his/her resulting mental representation. Four-year-olds have the

ability to establish this relationship between mental states, reality and actions. They can hold their own mental states and the other's mental states simultaneously in mind, hence 4-year-olds can differentiate false beliefs from true beliefs (Flavell, 1999; Perner, 1991).

There are different views in the literature on the relations between language and theory of mind. Some claim that developments in language abilities cause developments in Theory of Mind while others defend the opposite view (see Astington and Baird, 2005 for different approaches). What can be said, however, is that relations are reciprocal and change during development. In relation to children's understanding of evidential and epistemic markers in language, Matsui and her colleagues (2006) indicated that comprehension of evidential sentence-final particles was not associated with Theory of Mind capacity but comprehension of evidential verbs and comprehension of certainty particles and epistemic vocabulary were. In a similar vein, Aksu-Koç & Alıcı (2000) reported that children who passed the false belief tasks produced more epistemic vocabulary and more epistemic reasoning compared to children who did not. Aksu Koç (2009; Lucas et al., 2013) argued that early acquisition and use of evidential markers in language are likely to help children understand other minds, which is in turn likely to bring about advances in use of epistemic language related to degrees of certainty and reliability.

1.5 The present study: generalizability and its relation to source reliability

Information that is reliable is subject to generalization, whereas information that is unreliable is not likely to be generalized. As noted above, an investigation of when children understand that information obtained from a particular instance can be generalized to members of the same kind was carried out by Butler and Markman

(2012). These researchers examined the effect of pedagogical, accidental and intentional cues on inductive inferences and generalization with 3- and 4-year-olds using both ostensive demonstration and linguistic information. They attributed an inherent property (magnetism) to a novel object they called "blicket". In the pedagogical condition, they ostensively demonstrated this property on a single blicket with which they picked up paper clips while simultaneously saying "Look, the blicket is magnetic!". In the accidental condition, they accidentally dropped the "blicket" on the paperclips, saying "Oops!" and in the intentional condition they deliberately picked up paper clips with the blicket saying "Wow!", however, without establishing any joint attention. Then, in all three conditions the experimenter put 10 inert "blickets" on the table and let the child play with them for 60 seconds. Results demonstrated that English-speaking children as young as 4 years generalize the properties of novel objects that are linguistically presented and ostensively demonstrated for their benefit in pedagogical contexts more than when they are presented in contexts of intentional or accidental activity.

In a subsequent study, Tamm and her colleagues (2014) explored whether children would generalize in the absence of an ostensive demonstration of the property if the linguistic cue for generalizability is grammaticalized in the language, as it is in Turkish. They conducted a study with monolingual Turkish speaking 4- and 6-year-olds using Butler and Markman's (2012) methodology, however with the crucial difference that they conveyed the information that the blicket is magnetic only linguistically, without demonstrating its magnetic property. They had three conditions where they showed the blicket to the child and presented the sentence: (1) *Bilikit mıknatıslı-dır*, '(It is a fact that) blicket is magnetic' in the generic condition

(2) *Bilikit mıknatıslı-ymış*, '(I heard that- or -I infer that) blicket is magnetic' in the evidential condition

(3) *Bilikit mıknatıslı*, 'Blicket is magnetic' in the baseline condition.

Results showed that 4- and 6-year-olds generalize the property to members of the same kind in the generic *-DIr* condition but not in the evidential *-mIş* or the baseline conditions. This finding indicates that even 4 years old children can differentiate epistemic and evidential suffixes as cues indicating generic vs. non-generalizable knowledge.

Tamm et al.'s findings raise the question of why generalization is blocked when the evidential form is used. Is it because children interpret the *-mIş/-(y)mIş* inflection as marking information based on hearsay and therefore not reliable and generalizable? Or, is it because they interpret the *-mIş/-(y)mIş* inflection as expressing an inference based on a resultant state that constitutes only partial direct evidence and therefore not generalizable? Or is it because they interpret the *-mIş/-(y)mIş* inflection as expressing an inference marking a property specific to the particular object and therefore not generalizable?

The present study aimed to differentiate the inferential and hearsay functions of the evidential *-mIş/(y)mIş* to compare with the generic *-DIr* to see if children generalize differentially in response to the three different linguistic and situational cues. Thus, it was modeled after the Tamm et al. (2014) study except that the three different conditions were:

(1) *Bilikit mıknatıslı-dır*, '(It is a fact that) blicket is magnetic' in the generic condition

(2) *Bilikit mıknatıslı-ymış*, '(I infer that) blicket is magnetic' in the inferential condition

(3) *Bilikit mıknatıslı-ymış*, '(I heard that) blicket is magnetic' in the hearsay condition.

If children generalize more in the *-DIr* condition than either of the evidential conditions, this would confirm Tamm et al.'s results that they understand the generic meaning of the form. If they generalize more in the inferential *-mIş/(y)mIş* condition than in the hearsay *-mIş/(y)mIş* condition, this would mean that they treat inference from partially observable evidence as a more reliable source of information than hearsay which is secondhand information. If they generalize more in the hearsay condition than in the inference condition this may be because they treat secondhand information as more reliable. Therefore, the present study also explored whether children attribute different levels of reliability to the two uses of the *-mIş/-(y)mIş* form.

A further exploration into children's evaluations of reliability was made by assessing their Theory of Mind abilities, because if a child passes Theory of Mind tasks, s/he can correctly predict the knowledge access, false beliefs and actions of the other on the basis of the other's representation of reality which may be different from his / her own. This can be regarded as independent evidence that s/he understands the relation between types of information access and types of beliefs and can therefore base his / her judgments of reliability on the presence of direct vs. indirect information.

In this study, it was hypothesized that:

1. Children will generalize the property of an object to other instances of the same kind when the property is expressed by use of the generic marker *-DIr* more than when it is expressed by the inferential use of *-mIş/-(y)mIş* or the hearsay use of *-mIş/-(y)mIş*.

Children in the *-DIr* condition will test the magnetic quality of blickets (a) for a longer duration of time, (b) with more number of blickets, and (c) with more number of trials as compared to children in the inferential and hearsay conditions.

2. Children will generalize the property of an object to other instances of the same kind when it is conveyed by the inferential use of *-mIş/-(y)mIş* more than when it is conveyed by the hearsay use of *-mIş/-(y)mIş* because they will treat inferences based on partial direct evidence as more reliable than secondhand information.

   They will test the magnetic quality of blickets (a) for a longer duration of time, (b) with more number of blickets, and (c) with more number of trials in inferential use of *-mIş/-(y)mIş* condition than in hearsay use of *-mIş/-(y)mIş* condition.

3. Six-year-olds will show a higher level of understanding of *-DIr* as a generic marker than 4-year-olds do compared to inferential and hearsay *-mIş/-(y)mIş*.

   Six-year-olds will test the magnetic quality of blickets (a) for a longer duration of time, (b) with more number of blickets, and (c) with more number of trials than 4-year-olds do in *-DIr* condition compared to inferential and hearsay *-mIş/-(y)mIş* conditions.

4. Children will attribute higher degree of reliability to evidential information based on an inference from partial observable evidence, than to information based on a hearsay that indicates secondhand evidence.

5. Six-year-olds will attribute reliability to the inferential use of *-mIş/-(y)mIş* more than 4-year-olds do.

6. There will be a positive relationship between children's attributions of reliability and generalization patterns such that:

(i) In the inferential *–mIş/-(y)mIş* condition, children who attribute higher reliability to the inferential use of *-mIş* will show more generalization than children who attribute higher reliability to the hearsay use of *-mIş*.

(ii) In the hearsay *-mIş/-(y)mIş* condition, children who attribute higher reliability to the hearsay use of *-mIş* will show more generalization than children who attribute higher reliability to the inferential use of *-mIş*.

7. Children who have higher scores on Theory of Mind will attribute more reliability to information expressed with inferential *-mIş/-(y)mIş* than to information expressed with the hearsay *-mIş/-(y)mIş*.

CHAPTER 2

METHODOLOGY

2.1 Participants

A total of 102 participants were assessed on all tasks. However, after conducting experiments with the first 4 participants, one of the items of the Evidentiary Reliability Task had to be changed and a new item was added. Therefore, these 4 participants were excluded from the analyses. Another 2 participants were not interested in the tasks of the study and did not give codeable data, so they were also excluded from the analyses.

A total of 96 monolingual Turkish 4-year-olds (N = 48, 20 girls) and 6-year-olds (N = 48, 29 girls) constituted the final sample. Children were recruited from schools in upper middle-class residential areas in the two cities of Istanbul and Samsun. All of the participants whose parents gave consent to participate in the study were native Turkish speakers. Both parents were native speakers of Turkish but English language was also spoken in nine 4-year-olds' and five 6-year-olds' homes for some kind of activities such as reading a book or listening a music. Most of the 4 and 6 years old participants were exposed to English language at kindergarten or primary school. The average exposure is approximately 5 hours in a week.

Four-year olds were tested individually in their kindergartens and 6-year-olds who were attending the first grade were tested in their primary schools. The parental education and family income levels of the participants by age group are presented in Table 1. The distrinbution of the participants to the three experimental conditions by age (months) and sex are presented in Table 2.

Table 1. Parental Education and Income Levels of the Participants by Age Group

	Four-year-olds		Six-year-olds	
	Mean	SD	Mean	SD
Maternal Ed. (years)	15.65	2.99	16.13	2.74
Paternal Ed. (years)	15.77	3.12	16.57	2.72
Income (1000 TL)	8.88	4.41	8.40	3.13

Table 2. Distrinbution of the Participants to the Three Experimental Conditions by Age (Months) and Sex

	-DIR				Inferential *–mIş*				Hearsay *–mIş*			
	Female		Male		Female		Male		Female		Male	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
4-year-olds	53.14	4.25	51.33	3.04	55.00	3.36	52.67	4.00	51.44	3.94	51.14	2.79
6-year-olds	75.22	4.08	76.00	3.51	73.33	1.65	77.43	3.99	73.27	1.42	74.00	3.08

2.2 Procedure

Ethics approval was obtained from the Institutional Review Board for Human Subjects (İNAREK) of Boğazici University. A parental consent form and a demographic information questionnaire was sent to the parents and all the 4-year-old and 6-year-old children whose parents gave consent participated in the study. The

children were tested individually in a quiet room in their schools. The experiments were conducted only by the researcher herself. The testing session began with the Generalizability Task that approximately takes 10 minutes. This was followed by the administration of the Evidentiary Reliability Task that also approximately takes 10 minutes and Theory of Mind tasks (Unexpected Contents and Knowledge Access; (Keceli Kayisili & Acarlar, 2011) that together take 5 minutes. Finally, each child received TİFALDİ (Kazak Berument & Güven, 2010). The duration of TİFALDİ varied from 5 to 15 minutes depending on the child's knowledge of vocabulary. The whole session was video recorded. After the tasks were completed, participants were thanked and rewarded with two stickers for their participation, and their teachers were debriefed.

2.3 Measures

2.3.1 Generalizability task

This task is an adaptation of the "Blicket-Test" (Butler & Markman, 2012; Tamm et al., 2014) and was prepared to assess the comprehension of the generic function of the *-DIr* marker, in comparison with the inferential and hearsay uses of the *-mIş /-(y)mIş* marker. Eleven small, 2.5 cm x 2.5 cm x 5 cm rectangular wooden blocks (Butler & Markman, 2012) were used as novel objects. The blocks were called "bilikit" in Turkish, after "blicket" in English (Tamm et al., 2014). All wooden blocks were covered with green electrical tape for 2/3 of their length and with black electrical tape for 1/3 of their length. All blocks were perceptually identical but only one of them had a magnet hidden in its black end (the 'active block', the other ten blocks did not have a magnetic end (the 'inert blocks').

The children were assigned to three experimental conditions randomly. The procedure was the same in the three conditions except for the presentation of the stimulus sentence and for the context of the condition. All children were told that the aim is to teach them how to make a paper boat. First, they were introduced to a new toy, called "Bilikit". The experimenter asked the child to repeat the new toy's name, then put the blicket away to a place where the child could not reach. The experimenter then brought two boxes, each of which contained 1 blicket and 4 different distractor objects. She asked the child to show her the blicket to make sure that the child has learned the name of the novel object. Then, she put the boxes away and introduced the property of "being magnetic" using different everyday magnetic objects. The experimenter placed a box of paperclips and five different magnetic objects on the table and said "Look, these may all look different but they are all magnetic. Do you know what being magnetic means? For instance, they can stick on the refrigerator you have at home". She said "or you can pick these paperclips with the magnet", while picking up some paperclips with a magnetic object. She encouraged the child to try the magnetic objects by saying "Here, you can try them, too". Before beginning the task, it was made sure that children know what a "blicket" is and what being magnetic is.

Next, the experimenter put away the magnetic objects, and to assess children's interpretation of the different linguistic cues in terms of genericity, she took the "blikit" in her hand and presented the following stimulus sentences in the three different conditions:

1. Generic condition: *Bilikit mıknatıslıdır,* 'Blicket is magnetic [generally]'
2. Inferential condition: *Bilikit mıknatıslıymış,* 'Blicket is magnetic [evidently]'
3. Hearsay condition: *Bilikit mıknatıslıymış,* 'Blicket is magnetic [reportedly]'

In the generic condition, she took the blicket in her hand, pointed to the blicket and said *Bilikit mıknatıslıdır*, 'Blicket is magnetic [generally]'. In the inferential condition, she took the blicket, which already had a little magnetic object attached to its magnetic end, pointed to the blicket and said *Bilikit mıknatıslıymış*, 'Blicket is magnetic [evidently]'. In the hearsay condition, the cell phone of the experimenter rang. The experimenter pretended to answer the phone and looked at the blicket while she was talking. After she turned off the phone, she pointed to the blicket and said *Bilikit mıknatıslıymış*, 'Blicket is magnetic [reportedly]'. The phone call method was chosen to create the context where *–mIş* could be interpreted in its hearsay function, without presenting any source that implies credibility. In each condition, after the presentation of the stimulus sentence, the experimenter put away the blicket and said that she has to look for a piece of paper to make a paper boat. Before leaving the table, she pushed 10 inert blickets that are not magnetic but look exactly like the original blicket in front of the child close to the paperclips that were still on the table. "While I am looking for the paper to make the paper boat, you can play with these blickets". The child was free to play with these blickets for 60 seconds which were videotaped.

2.3.2 Evidentiary Reliability task

This task was designed to compare children's assessment of the levels of reliability implied by the *-mIş/-(y)mIş* marker used in its inferential versus hearsay functions. The task was prepared in Adobe Flash Program and presented on a 13 inch laptop computer. Children were introduced to the characters on a farm, some animals and a brother and a sister, Ali and Ece. A narrator's voice told a story about the characters and animals. The voice said "There is a lot of mischief going on in this farm, and it is

not well understood who does it. Cute monsters try to find out who did it. Let's see what is happening in this farm?". After the introduction, the experimenter asked all animals' names and the colors Ali's and Ece's dresses, which are the cues in some of the following animations. Before beginning the task, it was made sure that children knew the cues and characters well.

For each of seven trials (2 warm-up trials and 5 test trials), children saw a short animation film presenting scenes that depict a present resultant state from which the causal past process can be inferred. The scenes were described by use of the following 'change of state' verbs: *ye* 'eat', *ısır* 'bite', *patlat* 'pop', *dök* 'pour', *kır* 'break', *iç* 'drink', and *yırt* 'tear'. Each animation film showed an observable evidence for the change of state and two cute monsters, who have not witnessed the past process give contrasting information about it. While one of the monsters finds an observable evidence about who might have caused the change of state and declares information on that basis (inferential *-mIş/-(y)mIş* condition), the other monster does not see such evidence and reports information that he hears from another monster (hearsay *-mIş/-(y)mIş* condition). After the monsters give contrasting information, the participant was asked which information is correct. The cute monsters were counterbalanced according to whether the information was given by making an inference or on the basis of hearsay.

The task began with the warm-up phase where two items, one with *ye* 'eat' and the other with *ısır* 'bite' were always used. The experimenter showed the first warm-up item and told the child to listen carefully because she would ask some questions at the end of each story. In the first warm-up item encoded with the verb *yemek*, 'eat', first there was a whole cake on the table. Then, the cake was shown as missing a few slices and Ece's eyepatch was left on the table. A grey cute monster

came and asked who ate the cake. She did not see Ece's eyepatch because it was not seen on the screen from his point of view, and a red cute monster came and whispered to her ear. Then the first monster declared *Pastayı Ali yemiş,* '(I heard that) Ali ate the cake'. The experimenter asked the child *Pastayı kim yemiş? Peki, o canavar pastayı Ali'nin yediğini nasıl öğrendi? Birinden mi duydu? Bir şey mi gördü?* 'Who ate the cake? How did this monster find out Ali ate the cake? Did she hear from someone or did she see something?' in order to focus the child's attention to the source of knowledge. In the second part of this item, a blue cute monster came and saw Ece's eyepatch. This blue monster took the eyepatch and declared *Pastayı Ece yemiş,* '(I inferred that) Ece ate the cake'. Again, the experimenter asked the child *Pastayı kim yemiş? Peki, o canavar pastayı Ece'nin yediğini nasıl öğrendi? Birinden mi duydu? Bir şey mi gördü?* 'Who ate the cake? How did this monster find out Ece ate the cake? Did she hear from someone or did she see something?'. Then, the experimenter asked the child *Peki, sence pastayı kim yemiş?*, 'so, what do you think, who ate the cake?'. For the first 2 warm-up items, the experimenter asked how the informant found out who did the action. After the warm-up phase, 5 more animation videos which featured the verbs *patl*a 'pop', *dökmek*, 'pour', *kırmak*, 'break', *içmek* 'drink', and *yırtmak*, 'tear' were shown and the same procedure was repeated, however, without any questions about the source of knowledge. For instance, in one of items, there was a balloon and a few trees on the screen. Then, a popped balloon was presented on the screen with the feather of a bird. The purple cute monster came and asked who popped the balloon. This monster saw the feather, took it and declared *Balonu kuş patlatmış*, '(I inferred that) a bird popped the balloon'. The experimenter only asked *Balonu kim patlatmış?* 'Who popped the balloon?' to be sure that the child heard and understood the first source. Then, the

blue cute monster came but did not see the feather or any other evidence. The monster asked who popped the balloon, and the orange cute monster came and whispered to her ear. Then, the blue cute monster declared *Balonu sincap patlatmış*, 'I heard that- a squirrel popped the balloon'. Then, experimenter asked *Balonu kim patlatmış?* 'Who popped the balloon?' to be sure that the child heard and understood the second source. Lastly, the experimenter asked after the child heard the answers of both monsters *Peki, sence balonu kim patlatmış?* 'So what do you think, who popped the balloon?'.

The experimenter showed the items in random order. *Patla* 'pop' and *iç* 'drink' items began with the monster who was told the source; and *dök* 'pour' and *kır* 'break' items began with the monster who found a clue to make an inference about the source. *Yır*t 'tear' item had both of the versions, and half of the participants saw the hearsay first condition, while the other half saw the inference first condition to avoid primacy or recency effects. Also, the colors of the informant monsters were counterbalanced in terms of the source of knowledge. Example scenes with the characters are presented in appendix A (see Figure A1—A11).

2.3.3 Theory of Mind (ToM) tasks

Two Theory of Mind tasks, the Knowledge Access Task and the Unexpected Content Task were used. The Turkish protocols for both tasks were taken from Keceli Kayisili and Acarlar (2011). The English version of the protocol is presented in appendix B, and the Turkish version in appendix C.

2.3.3.1 The Knowledge Access task

This task originates from Pratt and Bryant (1990) and Pillow (1989), and a version adapted by Wellman & Liu (2004) was used in the present study. The task aims to measure the ability to hold one's own knowledge and the other person's knowledge simultaneously in mind (Wellman & Liu, 2004). The experimenter showed the child a box and asked *Bu bir kutu. Sence bu kutunun içinde ne var?* "Here is a box. What do you think is inside of this box?". After the child's answer, the experimenter said *Evet güzel bir tahmin. Hadi açalım bakalım içinde ne var!* "That is a good guess, let's open and see what it is inside!" and opened the box and said *Aa bak, içinde bir köpek var!* "Look, it is a toy dog inside!". Then, she put the dog back in the box and asked the child *Kutuda ne var?* "What is in the box?" as first control question. After the child's answer, the experimenter presented a doll and said *Bu çocuk daha once bu kutuyu hiç görmedi ve hiç açmadı. Bu çocuk kutunun içine baktı mı? "*This boy has not seen the box and has not opened it. Did the boy look inside of the box?" as second control question. Then, the experimenter asked *Peki, bu çocuk kutunun içinde ne olduğunu biliyor mu?* "Does the boy know what is inside of the box?" as a target question. After the child's answer, the experimenter asked "Why?". Successful performance on the task shows that the child understands the relation between mode of access to information and resulting knowledge state, that is, between source and knowledge.

2.3.3.2 The Unexpected Contents task

The task is modeled on Hogrefe, Wimmer and Perner (1986), as one of the standard false belief tasks. Children were shown a candy box (Bonibon) that all children are familiar with and asked *Bak, şimdi sana bir şey göstereceğim. Sence bu kutunun*

içinde ne var? "Look! I will show you something. What do you think is inside this box?" After the child's answer, which is expected to be "candy", the experimenter opened the box and showed that there were pencils in the box. She said *Aa kutuda ne varmış?* "What is in the box?". After the child's answer that there are pencils in the box, the experimenter said *Evet, kutunun içinde kalemler varmış. Peki, ben bu kutuyu açmadan once sen içinde ne olduğunu sanmıştın?* "Yes, there are pencils inside of the box. What did you think was in the box before I opened it?" as first target question. If the child answered "Candies!", this indicated that s/he can keep track of his own changing mental representations. Then, the experimenter said *Arkadaşın henüz bu kutuyu ve içindekileri göremedi, birazdan onu da çağıracağım. Arkadaşın, ben bu kutuyu açmadan önce, içinde ne olduğunu sanır?* "Your friend hasn't seen this box or what is inside of it, yet. I will call her/him shortly. What would s/he think is in the box before I open it?" as the second target question. The correct answer, "candies" indicates that the child is able to represent the false belief of someone who has not seen the real contents of the box. After the child's answer, the experimenter asked why s/he thought like that.

2.3.4 Turkish Expressive and Receptive Language (TİFALDİ) test

This test which assesses both the receptive and expressive vocabulary skills of Turkish children is modelled after the PPVT and has been developed for Turkish based on a representative sample of 3755 children aged 2-13 (Kazak Berument & Güven, 2010). It is a valid and reliable test to assess the vocabulary skills of Turkish children, Cronbach's alpha = .99 per 104 items, Cronbach's alpha calculated separately for each age group (2-12 year-olds) changes between .88 and .96 per the group of items. The test has 104 concrete and abstract Turkish words selected from compre-

hensive Turkish dictionary and word frequency list that are suitable for assessing the vocabulary skills of 2-12 years-old children. In this test, the experimenter presented a word that is appropriate for the child's age group and asked the child to point to the correct picture corresponding to the meaning of the word among four pictures. The child's task was to identify the picture that represents the word's meaning. The test was terminated when child made 8 mistakes within the last 10 items. Some example pictures from the list of words are presented in appendix D (see figure D1 and figure D2).

2.3.5 Demographic information questionnaire

A demographic information questionnaire was completed by each participant's parents. This questionnaire asked for information about the child's age and sex, parents' education levels and occupations, languages spoken at home, and income level (See Appendix E for the demographic information questionnaire in English and Appendix F in Turkish).

2.3.6 Computer

A MacBook Air 13" DC i5 1.6GHz 8GB 128GBflash with Adobe Flash Player Program was used in the study.

2.3.7 Coding and inter-rater reliability

The data were coded by the researcher. Coding protocols are explained for each task in this section. For reliability purposes a trained coder coded 25% of the data. The trained coder randomly chose 25 participants' videos for each task. Then, inter-rater

reliability analyses were run for each task. The reliabilites are reported below for each task.

2.3.7.1 Generalizability task scoring

Five types of scores were calculated on the basis of the video records. Three scores which were duration, number of blickets and total number of trials were used as dependent variables in the analyses (Tamm et al., 2014).

i. Duration: the time spent on trying the inert blickets for magnetism during 60 seconds.

ii. Number of blickets: the number of inert blickets explored for magnetism, to see if it has the magnetic property.

iii. Number of trials: the number of times the child tests inert blickets.

iv. Each other trials: the number of times the child tested two inert blickets by making the (supposed) magnetic parts of each touch one another.

v. Total number of trials: the sum of "number of trials" and "each other trials"

Inter-rater reliability analyses were run for each variable. Kappa is .99 for "duration", .90 for the "number of trials", 1.00 for the "number of blickets", .94 for the number of "each other trials", .91 for the "total number of trials".

2.3.7.2 Evidentiary Reliability task scoring

The number of times the participant chose as a reliable informant the monster that uses a *-mIş/-(y)mIş* statement to express an inference from observable evidence in context constituted the 'inference reliability score' and the number of times the participant chose as a reliable informant the monster that uses a *-mIş/-(y)mIş* statement on the basis of what it heard from another monster constituted the 'hearsay

reliability score'. Participants' choices were scored as 1 point for each item. At the end, each participant was assigned a continuous score between 0-5 points. If a child chose the inferential source 3 times and hearsay source 2 times out of 5 items, his/her 'inference reliability score' was 3 and 'hearsay reliability score' was 2. In addition, a "reliability preference" variable was composed as a categorical variable. If inference reliability score was higher than hearsay reliability score, inference has been attributed greater reliability and if hearsay reliability score was higher than inference reliability score, hearsay has been attributed more reliability. Participants' choices were scored as 1 for inference and 0 for hearsay. That is, if a child's inference reliability score was 3 and hearsay reliability score 2, her/his reliability preference was codded as 1, and if a child's inference reliability score was 2 and hearsay reliability score 3, her/his reliability preference was coded as 0. Children were assigned to one or the other category according to their choices.

Inter-rater reliability analyses were conducted for each variable. There was no disagreement between coders for any of the items (*patl*a 'pop', *dökmek*, 'pour', *kırmak*, 'break', *içmek* 'drink', and *yırtmak*, 'tear'). Thus Kappa is 1.00 for "inference reliability" and "hearsay reliability" scores.

2.3.7.3 Theory of Mind (ToM) scoring

The scores for Knowledge Access and Unexpected Contents tasks were calculated separately. The ToM score was formed by summing Knowledge Access and Unexpected Contents task scores.

Knowledge Access task: Children were given a score of 1, if they answered "No" in response to the question "Does the boy know what is inside of the box?"; and a score

of 1, if they gave a correct explanation to the question “Why?”. The maximum score to be obtained was 2 for this task.

Unexpected Contents task: Children were given a score of 1 if they answered “candy” in response to the question “What did you think was in the box before I opened it?”, and a score of 1 if they answered “candy” in response to the question “Your friend hasn’t seen this box or what is inside of it, yet. I will call him / her shortly. What would s/he think is in the box before I open it?”, and a score of 1 if they gave a correct explanation to the question “Why?”. The maximum score to be obtained was 3 for this task. The maximum ToM score to be obtained was 5.

Inter-rater reliability analyses were run for each variable. Kappa is 1.00 for the Knowledge Access scores and .95 for the Unexpected Contents scores.

2.3.7.3 Turkish Expressive and Receptive Language (TİFALDİ) test scoring

A raw score, a standart score, an equivalent age and a percentile rank were calculated for each child according to the scoring instructions of the test (Kazak-Berument & Güven, 2010). These four variables were calculated and coded in SPSS. TİFALDİ raw score which is affected by child’s age, and TİFALDİ Standart Score which eliminates the age effect were used for the analyses.

Inter-rater reliability analyses were run for each variable. Kappa is .92 for the “TİFALDİ Raw Score”, .87 for the “TİFALDİ Standart Score”, .92 for the “equivalent age”, and .82 for the “percentile rank”.

CHAPTER 3

RESULTS

3.1 Preliminary analyses

3.1.1 Descriptive statistics

Before the main analyses, all variables were checked to see whether the sample is normally distributed. In the Generalizability task, skewness and kurtosis values were not within acceptable limits (+- 1.96 SD) except for the duration variable for which the kurtosis value was -.68 (SE = .49). For duration, skewness was .84 (SE = .25), for number of blickets, skewness was 2.15 (SE = .25) and kurtosis 5.97 (SE = .49), and for total trials variable, skewness was 1.70 (SE = .25) and kurtosis 2.48 (SE = .49). Therefore, logarithmic transformations were applied for duration, number of blickets and total number of trials variables. These transformed scores were used in the analyses. Descriptive statistics (pre-transformation) for the main variables are shown in Table 3.

3.1.2 Analyses for gender

Preliminary analyses were conducted to see whether there is a gender effect on children's performance on each task. First, a one-way MANOVA was conducted to see the effect of gender (2) on generalization variables (3) of the Generalizability task, that is, duration, number of blickets and total number of trials. The MANOVA revealed no significant effect of gender on generalization, *Wilks' Lambda (Λ)* = .99, $F = .37$, $p > .05$. Test of between-subjects effects conducted to see the effect of gender on each dependent variable separately showed no significant effect on duration, $F\ (1, 94) = .57$, $p > .05$, on number of blickets, $F\ (1, 94) = 1.05$, $p > .05$, or

on total number of trials, $F(1, 94) = .49$, $p > .05$. Therefore, gender was not investigated as an independent variable in the further analyses of generalization.

Table 3. Descriptive Statistics (Number of Participants, Minimum and Maximum Values, Means and Standard Deviations) for the Main Dependent Variables by Age

Age	4-Year-Olds					6-Year-Olds				
Variable	N	Min.	Max.	Mean	SD	N	Min.	Max.	Mean	SD
Duration	48	0	60	14.60	3.65	48	0	60	25.58	19.99
Number of Blickets	48	0	10	1.27	1.85	48	0	10	2.29	2.05
Total Number of Trials	48	0	33	4.65	8.28	48	0	39	8.94	8.83
Inference Preference	48	1	5	3.19	.98	48	2	5	4.02	1.13
Hearsay Preference	48	0	4	1.81	.98	48	0	3	.96	1.13
Knowledge Access	48	0	2	1.50	.72	48	0	2	1.92	.40
Unexpected Contents	48	0	3	1.67	1.14	48	0	3	2.13	1.00
ToM Score	48	0	5	3.17	1.60	48	0	5	4.04	1.13
TİFALDİ Raw Score	48	24	89	61.54	17.35	48	60	100	86.08	8.04
TİFALDİ Standart Score	48	90	137	120.8	12.90	48	99	140	123.69	9.81

In order to explore whether there is an effect of gender (2) on reliability (2), a one-way MANOVA was conducted. The MANOVA indicated that there is no significant effect of gender on reliability, *Wilks' Lambda (Λ)* = .98, $F = 1.11$, $p > .05$. Tests of between-subjects effects demonstrated that gender has no significant effect on inference reliability score, $F(1, 94) = 1.34$, $p > .05$; or on hearsay reliability score, $F(1, 94) = 1.54$, $p > .05$. Therefore, gender was not investigated as an independent variable in the further analyses of reliability.

Finally, in order to see whether there is an effect of gender on Theory of Mind performance and on children's general linguistic competence seperate analyses were conducted. Results of the one-way ANOVA on ToM scores revealed a marginally significant trend for gender, girls had higher scores than boys on total ToM scores, $F(1, 92) = 3.55$, $p = .06$, $\eta p^2 = .04$. Results of the one way ANOVA on TİFALDİ raw scores showed that gender did not significantly affect TİFALDİ raw scores, $F(1, 92) = .03$, $p > .05$. Therefore, gender was not explored as an independent variable in further TİFALDİ analyses but it was included in ToM analyses.

3.1.3 Analyses for general language competence

Since this is a psycholinguistic study where the independent variables are linguistic, a one way ANOVA was conducted to see whether children have age appropriate language competencies as indexed by vocabulary. The means and standard deviations of TİFALDİ Raw Scores and TİFALDİ Standard Scores are presented in Table 4 by age and gender.

Table 4. Means and Standart Deviations of TİFALDİ Standard Scores and Raw Scores by Age and Gender

Gender	Age Group	TİFALDİ Standard Score			TİFALDİ Raw Score		
		N	Mean	SD	N	Mean	SD
Female	4-year-olds	20	121.35	11.41	20	61.70	15.92
	6-year-olds	29	123.59	8.30	29	85.86	5.92
Male	4-year-olds	28	120.36	14.05	28	61.43	18.59
	6-year-olds	19	123.84	12.00	19	86.42	10.67

Results of the one way ANOVA indicated that 6-year-olds have higher TİFALDİ raw score ($N = 48$, $M = 86.08$, $SD = 8.04$) than 4-year-olds ($N = 48$, $M = 61.54$, $SD = 17.35$), $F(1, 94) = 79.07$, $p < .001$, $\eta p^2 = .46$.

According to TİFALDİ (receptive) age norms (Kazak-Berument & Güven, 2010), the raw score of 61, which is the 4-year-olds' mean score in our study, is equivalent to the vocabulary score of 6;02 years old children and the raw score of 86, which is the 6-year-olds' mean score in our study, is equivalent to the vocabulary score of 8;06 years old children. It is clearly observed that our participants have a general language competencies indexed by vocabulary well above their age level.

3.2 Main analyses

The present study aimed to answer the following questions: (1) whether children generalize the property of an object to other instance of the same kind when the property is expressed by use of the generic marker *-DIr* more than when it is

expressed by the inferential or the hearsay use of *-mIş/-(y)mIş*, (2) whether the level of understanding of *-DIr* as a generic marker increases with age, (3) whether children generalize the property of an object to other instances of the same kind when it is conveyed by the inferential use of *-mIş/-(y)mIş* more than when it is conveyed by the hearsay use of *-mIş/-(y)mIş,* (4) whether children attribute higher degree of reliability to evidential information based on an inference from partial observable evidence, than to information based on an hearsay that indicates indirect secondhand evidence, (5) whether attributing reliability to the inferential use of *-mIş/-(y)mIş* increases by age, (6) whether there is a positive relationship between children's attributions of reliability to inferential / hearsay sources and the generalization patterns in terms of inferential / hearsay patterns, (7) whether there is a positive relationship between ToM scores and attributing higher reliability to information expressed with inferential *-mIş/-(y)mIş* than to information expressed with the hearsay *-mIş/-(y)mIş.*

3.2.1 The effect of generic *-DIr*, inferential *-mIş/-(y)mIş* and hearsay *-mIş/-(y)mIş* markers on generalization

In order to test the hypothesis that children will generalize the property of an object to other instance of the same kind when the property is expressed by the use of the generic marker *-DIr* more than when it is expressed by the inferential or the hearsay uses of *-mIş/-(y)mIş,* a two-way MANOVA was conducted to examine the effect of age (2) x linguistic markers (3) on the three generalization variables of duration, number of blickets and total number of trials.

Logarithmic transformed scores were used for the dependent variables since they were not normally distributed. After this transformation, Levene's test of equality of error variances was not significant, $p > .05$. However, Box's M test

indicated the violation of the assumption of homogeneity of variance-covariance matrices, $p < .001$. Since there are a reasonable number of participants and our sample sizes are equal in each group a MANOVA can still be conducted as a valid measure. The two-way MANOVA revealed a significant effect of age, *Wilks' Lambda (Λ)* = .78, $F = 8.09$, $p < .001$, $\eta p^2 = .22$ and a significant effect of linguistic markers, *Wilks' Lambda (Λ)* = .82, $F = 3.04$, $p < .01$, $\eta p^2 = .10$ on generalization. There was no significant interaction between age and linguistic markers, *Wilks' Lambda (Λ)* = .92, $F = 1.31$, $p > .05$.

Test of between-subject effects was run to explore the effect of independent variables on each dependent variable separately. The results indicated that the effect of age on generalization was significant, as 6-year-olds tried blickets for longer duration of time ($M = 1.24$, $SD = .08$) than 4-year-olds ($M = .67$, $SD = .08$) to see if it had the magnetic property, $F(1, 90) = 24.69$, $p < .001$, $\eta p^2 = .22$ (See Figure 1). Six-year-olds also tried more number of blickets ($M = .46$, $SD = .04$) than 4-year-olds ($M = .25$, $SD = .04$), $F(1, 90) = 17.27$, $p < .001$, $\eta p^2 = .16$, and made more trials ($M = .82$, $SD = .06$) than 4-year-olds ($M = .42$, $SD = .06$) to test for the magnetic property of the blickets, $F(1, 90) = 20.74$, $p < .001$, $\eta p^2 = .19$, (See Figure 2 and Figure 3, respectively).

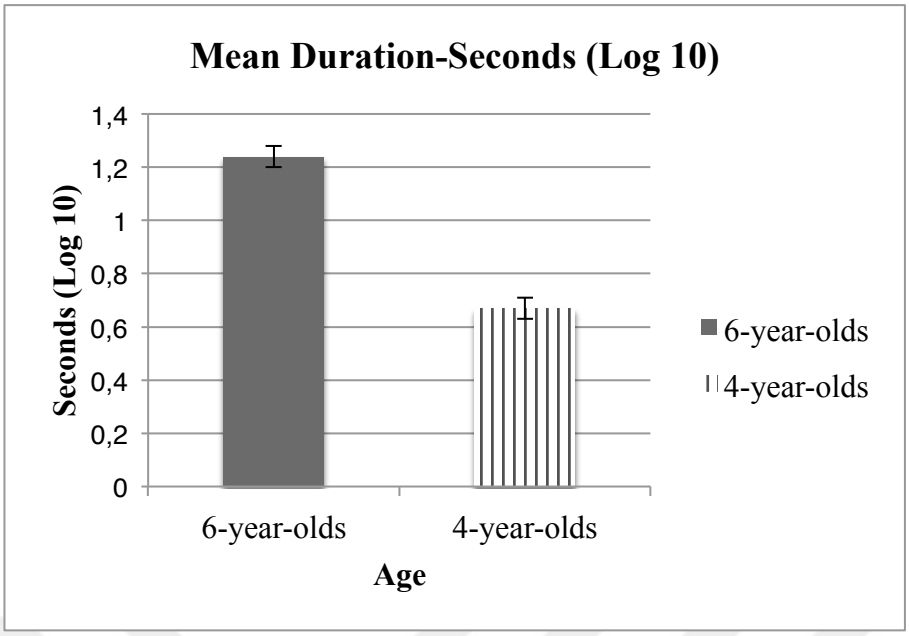

Figure 1. Mean duration of testing the blickets for magnetic property by age on the Generalizability task regardless of linguistic marker condition

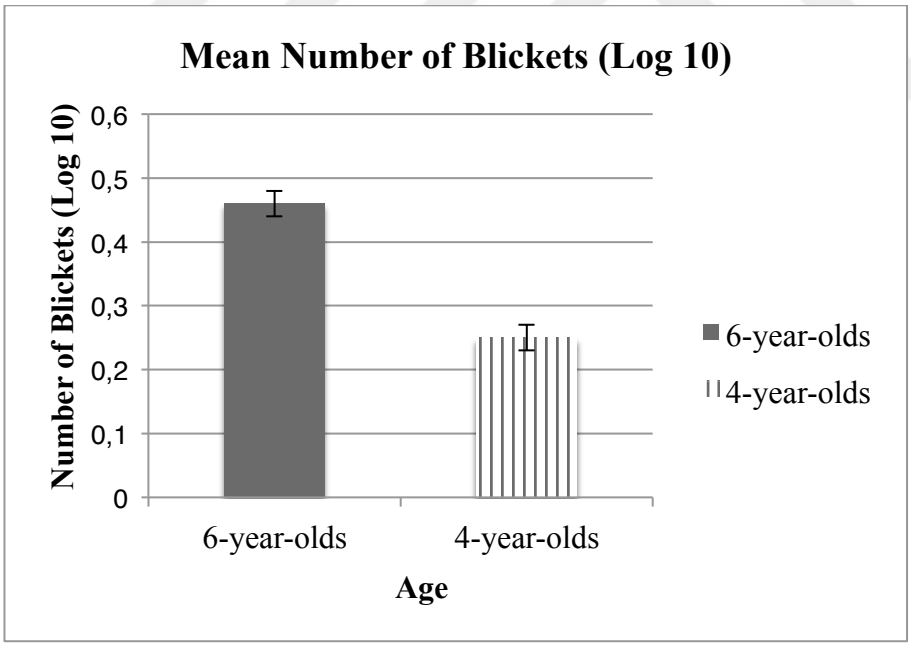

Figure 2. Mean number of blickets tested for magnetic property by age on the Generalizability task regardless of linguistic marker condition

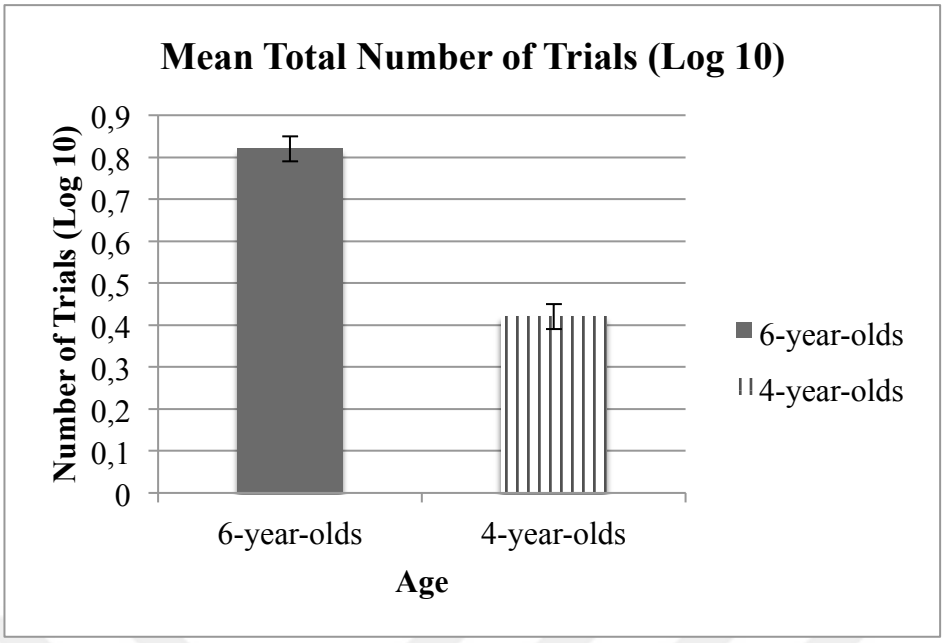

Figure 3. Mean number of total trials carried out for magnetic property by age on the Generalizability task regardless of linguistic marker condition

Test of between-subjects effects also showed significant effects of linguistic markers on duration of time children tried the blickets $F(2, 90) = 9.09$, $p < .001$, $\eta p^2 = .17$, on the number of blickets they tried, $F(2, 90) = 7.14$, $p < .01$, $\eta p^2 = .14$, and on the number of trials they carried out $F(2, 90) = 9.16$, $p < .001$, $\eta p^2 = .17$ to test the magnetic property of the blickets.

To test for the mean differences between linguistic conditions Bonferroni corrections were applied. The comparisons demonstrated that children tried blickets for longer duration of time in the *-DIr* condition *(M* = 1.30*, SD* = .59) than the inferential *-mIş/-(y)mIş* condition *(M* = .79*, SD* = .68), $p < .01$, and the hearsay *-mIş/-(y)mIş* condition *(M* = .78*, SD* = .65), $p = .001$. Children also tried more number of blickets in the *-DIr* condition *(M* = .48*, SD* = .28) than the inferential *-mIş/-(y)mIş* condition *(M* = .30*, SD* = .29), $p < .01$ and the hearsay *-mIş/-(y)mIş* condition *(M* = .28*, SD* = .23), $p < .01$. The total number of trials were also higher

in the *–DIr* condition *(M* = .88, *SD* = .49) than the inferential *-mIş/-(y)mIş* condition *(M* = .50*, SD* = .48), $p < .01$ and the hearsay *-mIş/-(y)mIş* condition *(M* = .48*, SD* = .45), $p < .01$. However, these comparisons showed that there is no significant mean difference between inferential *-mIş/-(y)mIş* and hearsay *-mIş/-(y)mIş* conditions on duration, number of blickets or total number of trials variables, $p > .05$ (See Figure 4, Figure 5 and Figure 6, respectively).

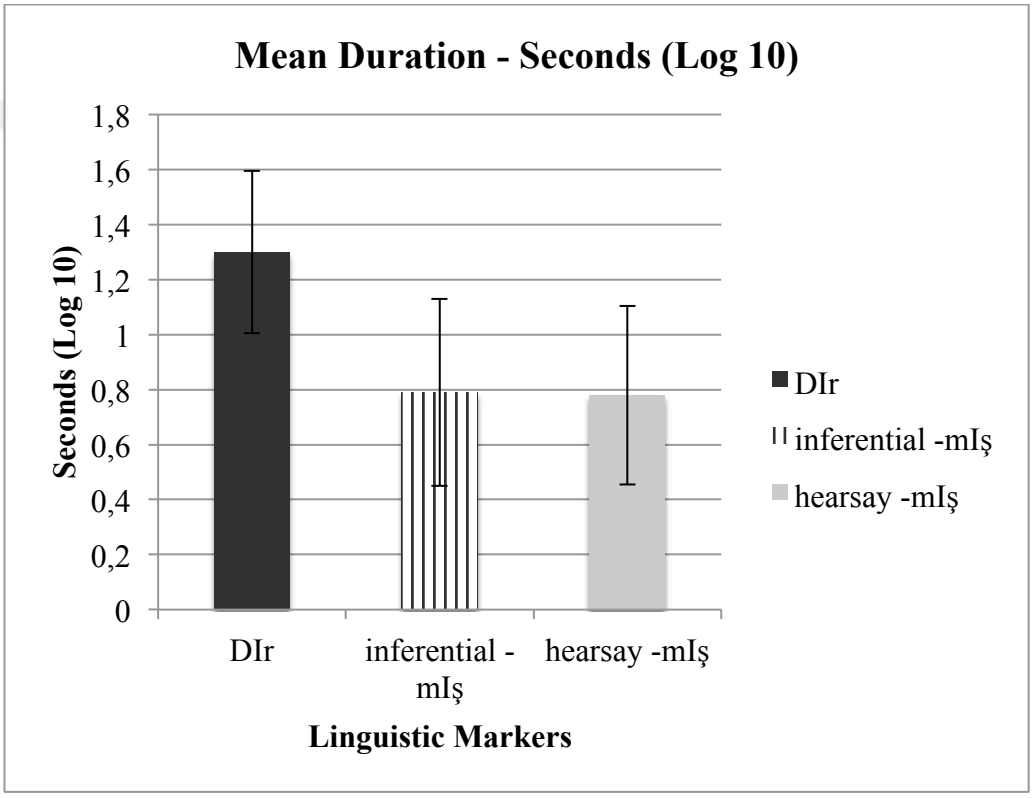

Figure 4. Mean duration of testing the blickets for magnetic property by linguistic markers on the Generalizability task regardless of age

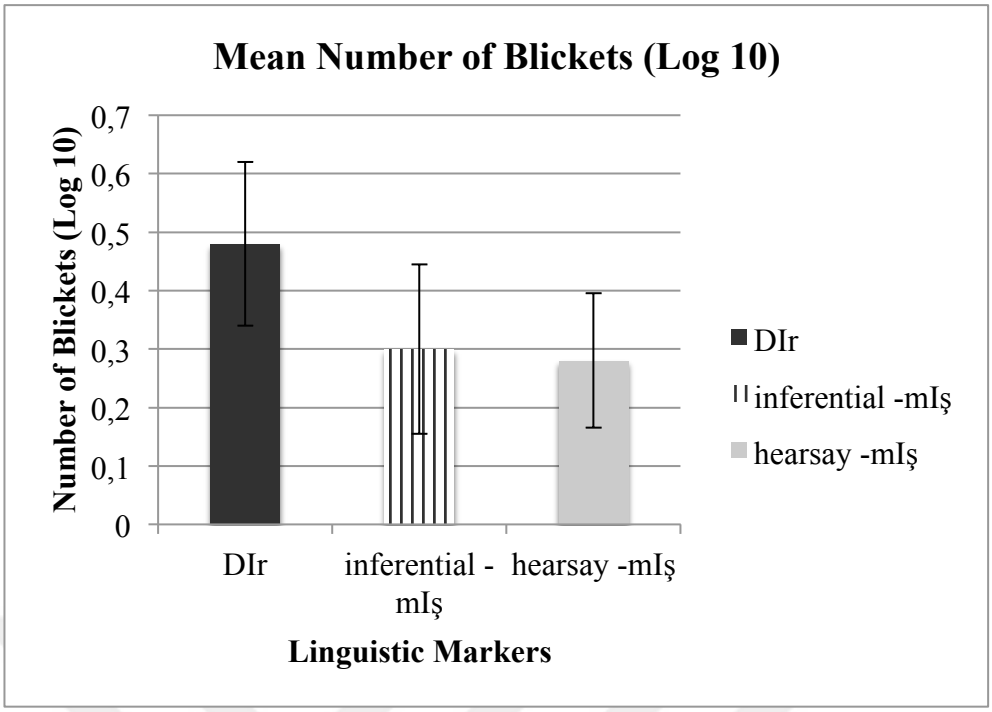

Figure 5. Mean number of blickets tested for magnetic property by linguistic markers on the Generalizability task regardless of age

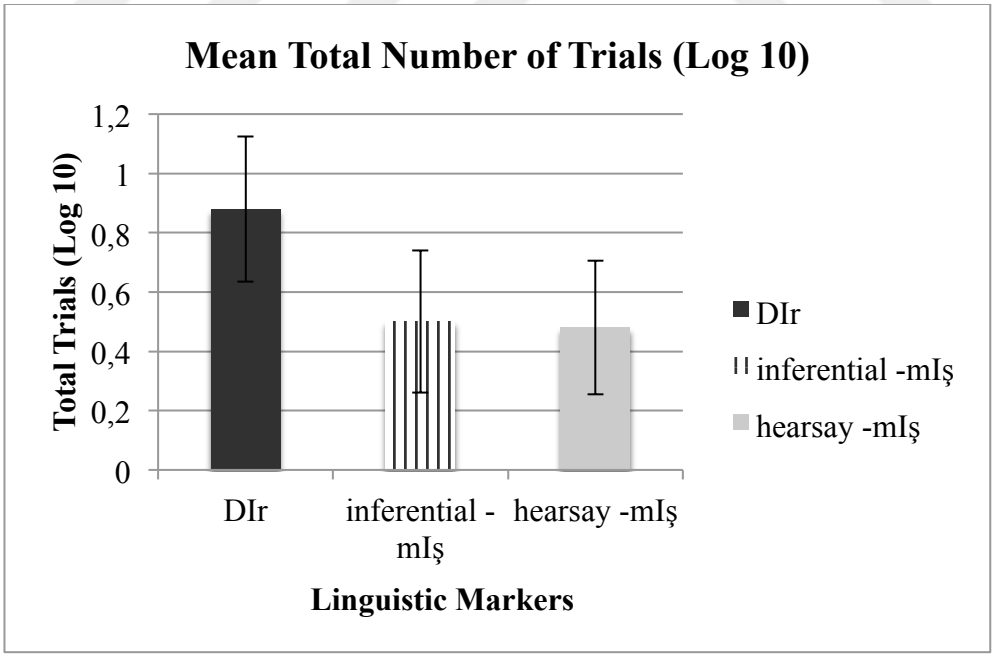

Figure 6. Mean number of total trials carried out for magnetic property by linguistic markers on the Generalizability task regardless of age

Test of between-subjects effects demonstrated that there is a significant interaction between age and linguistic markers on the duration variable, $F(2, 90) = 3.89$, $p = .025$, $\eta p^2 = .08$, and a marginally significant interaction between age and linguistic markers on total number of trials variable, $F(2, 90) = 2.99$, $p = .055$, $\eta p^2 = .06$. However, there is no interaction between age and linguistic markers on number of blickets variable, $F(2, 90) = 1.92$, $p > .05$.

To examine the source of the interaction and to see the effects of linguistic markers on generalizability in terms of duration, number of blickets and total number of trials for 4-year-olds and 6-year-olds separately, a one-way MANOVA was conducted for each age group. The MANOVA indicated a significant effect of linguistic markers on generalization for 4-year-olds, *Wilks' Lambda (Λ)* = .65, $F = 3.50$, $p < .01$, $\eta p^2 = .20$. However, there was no significant effect of linguistic markers on generalization for 6-year-olds, *Wilks' Lambda (Λ)* = .94, $F = .48$, $p > .05$.

Test of between-subjects for 4-year-olds showed significant differences between linguistic conditions in terms of the variables of duration of time spent for testing the blickets for magnetic property, $F(2, 45) = 9.91$, $p < .001$, $\eta p^2 = .31$; in terms of the number of blickets tried, $F(2, 45) = 6.86$, $p < .01$, $\eta p^2 = .23$, and also in terms of the number of trials carried out, $F(2, 45) = 11.31$, $p < .001$, $\eta p^2 = .34$. However, 6-year-olds did not show any significant difference between linguistic conditions for generalization, either in terms of duration $F(2, 45) = 0.88$, $p > .05$, or number of blickets $F(2, 45) = 1.40$, $p > .05$; or number of trials $F(2, 45) = 0.87$, $p > .05$. This result did not support our third hypothesis that 6-year-olds will show a higher level of understanding of *-DIr* as a generic marker than 4-year-olds as compared to inferential use of *-mIş/-(y)mIş* and hearsay use of *-mIş/-(y)mIş*.

Results of Bonferroni multiple comparisons indicated that 4-year-olds generalized the property of an object to other instance of the same kind when the property is expressed by use of the generic marker *-DIr* more than when it is expressed by the inferential use of *-mIş/-(y)mIş,* and more than the hearsay use of *-mIş/-(y)mIş* in terms of duration, $p < .01$, number of blickets, $p < .01$, and number of trials, $p < .01$. However, the mean comparisons showed that there is no significant mean difference between inferential *-mIş/-(y)mIş* and hearsay *-mIş/-(y)mIş* conditions on duration, number of blickets or total number of trials variables, $p > .05$ (see Table 5 for the means and standart deviations, and Figure 7, Figure 8 and Figure 9 for the histograms).

Table 5. Means and Standard Deviations for Duration, Number of Blickets and Total Number of Trials by Age Group and Linguistic Marker Condition

Age Groups	Linguistic Markers	Duration (Seconds-Log10)		Number of Blickets (Log10)		Total Number of Trials (Log10)	
		M	SD	M	SD	M	SD
4 year olds	Generic *-DIr*	1.24	.68	.45	.29	.83	.53
	Inferential *-mIş*	.37	.59	.15	.25	.20	.38
	Hearsay *-mIş*	.42	.59	.16	.22	.24	.33
6 year olds	Generic *-DIr*	1.36	.50	.52	.25	.93	.47
	Inferential *-mIş*	1.22	.46	.46	.24	.80	.39
	Hearsay *-mIş*	1.14	.50	.39	.16	.72	.42

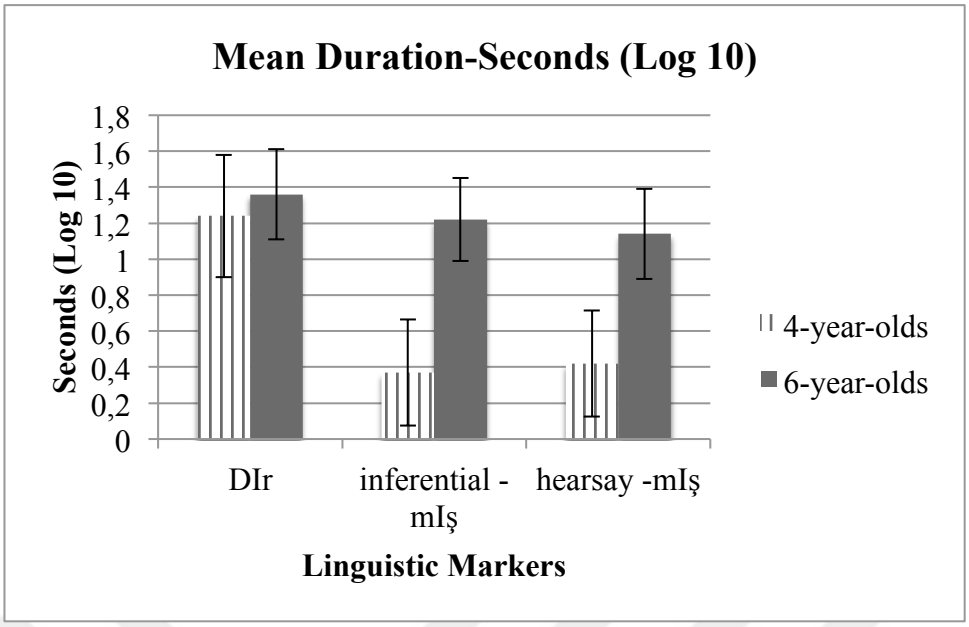

Figure 7. Mean duration of testing the blickets for magnetic property by age and linguistic markers on the Generalizability task

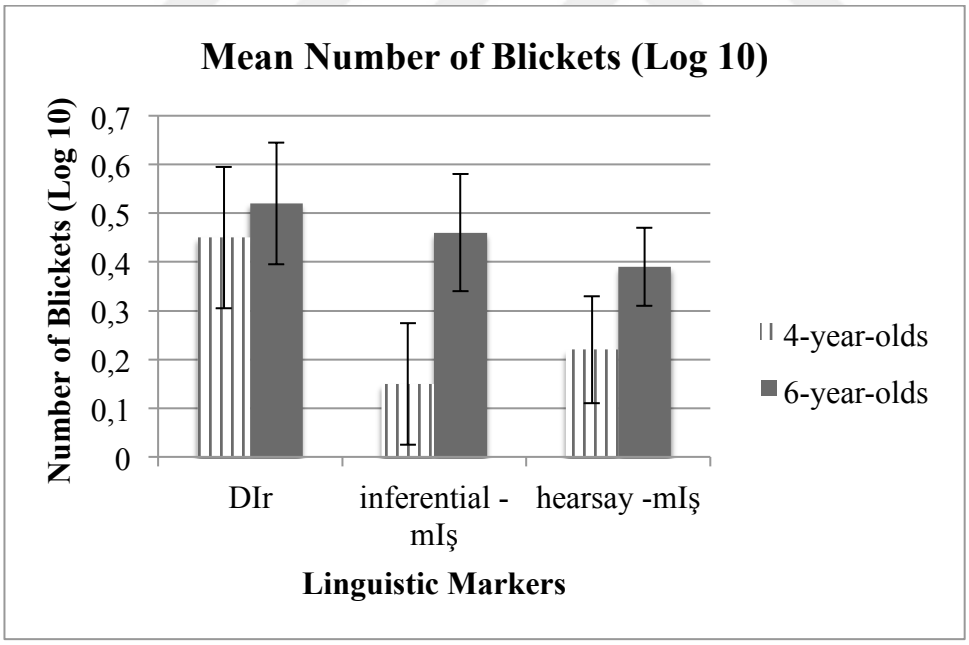

Figure 8. Mean number of blickets tested for magnetic property by age and linguistic markers on the Generalizability task

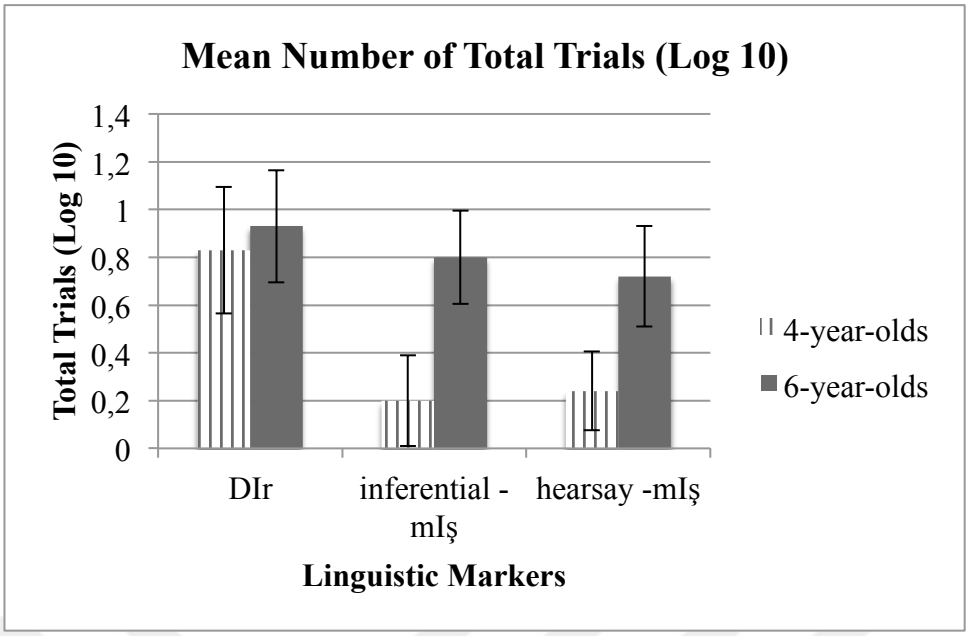

Figure 9. Mean number of total trials carried out for magnetic property by age and linguistic markers on the Generalizability task

3.2.2 The effect of inference vs. hearsay as evidential source on judgements of reliability of information

The total score of a child's choice of inference as the more reliable source (maximum score = 5) constituted the inference reliability score. In order to test the hypothesis that children attribute higher degree of reliability to information based on an inference from partial observable evidence than to information based on a secondhand evidence, a one sample t-test was conducted for 4-year-olds and 6-year-olds separately. Results showed that both 4 years old and 6 years old children chose inference as a more reliable source than hearsay above chance level, $t(47) = 22.49$, $p < .001$; and $t(47) = 22.87$, $p < .001$, respectively.

An independent samples t-test was run to test the hypothesis that 6-year-olds attribute reliability to the inferential use of *-mIş/-(y)mIş* more than 4-year-olds do. The results showed that 6-year-olds chose inference as the more reliable source more

than 4-year-olds did, $t(94) = -3.88$, $p < .001$. This finding indicates that children's understanding of a reliability difference between the inferential and hearsay uses of *–mIş/-(y)mIş* increases across age groups (See Table 6 for means and standard deviations of children's reliability attributions, and Figure 10 for the historgrams).

Table 6. Means and Standard Deviations of Children's Reliability Attributions

Age Groups		N	Mean	SD
4-year-olds	Inference Reliability Score	48	3.19	.98
	Hearsay Reliability Score	48	1.81	.98
6-year-olds	Inference Reliability Score	48	4.02	1.12
	Hearsay Reliability Score	48	.96	1.13

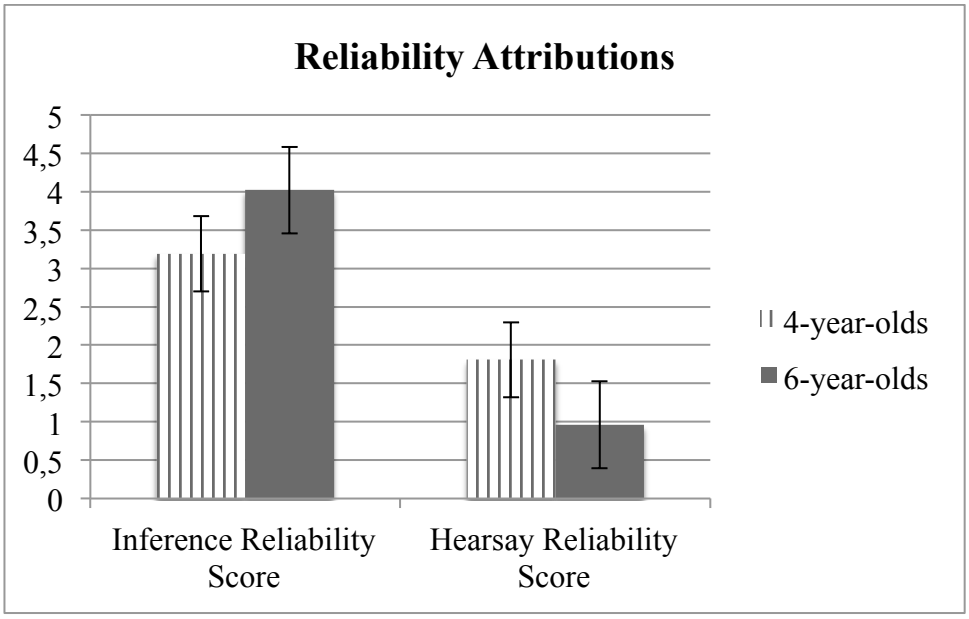

Figure 10. Reliability attiributions of children by age

3.2.3 Relationship between generalizability and reliability

In order to explore whether there is a relationship between children's attributions of reliability and their generalization patterns, a two-way MANOVA was conducted to examine the effect of reliability preference (2) and linguistic markers (3) on the generalization variables of duration, number of blickets and total number of trials. Logarithmic transformed scores were used for the dependent variables.

The analysis did not reveal a significant effect of either reliability preference, *Wilks' Lambda (Λ)* = .95, $F = 1.65$, $p > .05$, or a significant effect of linguistic markers, *Wilks' Lambda (Λ)* = .92, $F = 1.25$, $p > .05$. There was no interaction between reliability preference and linguistic markers, *Wilks' Lambda (Λ)* = .94, $F = .86$, $p > .05$.

Tests of between-subjects effects showed that the effect of reliability preference on generalizability in terms of duration was significant, suggesting that children, who chose inferential source as more reliable, tried blickets for longer duration of time for magnetism ($M = 1.03$, $SD = .07$) than children who chose hearsay as the more reliable source ($M = .69$, $SD = .14$), $F(1, 90) = 4.49$, $p < .05$, $\eta p^2 = .05$. Children who chose inferential source as more reliable also tried blickets with more number of trials ($M = .67$, $SD = .05$) than children who chose hearsay source as more reliable ($M = .42$, $SD = .11$), $F(1, 90) = 4.55$, $p < .05$, $\eta p^2 = .05$. However, there was no significant main effect of reliability preference on generalizability in terms of the number of blickets children tried to explore for magnetism, $F(1, 90) = 2.25$, $p > .05$.

Tests of between-subjects effects also indicated that there was a significant main effect of linguistic markers on generalizability in terms of duration, $F(2, 90) = 3.34$, $p < .05$, $\eta p^2 = .07$, and total number of trials, $F(2, 90) = 3.21$, $p < .05$, $\eta p^2 = .07$,

but not on the of number of blickets children tried to explore for magnetism, $F(1, 90) = 1.88$, $p > .05$. There was no significant interaction between reliability preference and linguistic markers on duration, $F(2, 90) = 1.71$, $p > .05$, on number of blickets, $F(2, 90) = 1.71$, $p > .05$ and, on total number of trials, $F(2, 90) = 1.39$, $p > .05$.

Follow up Bonferroni pairwise comparisons indicated that children tried blickets for longer duration of time when the property was expressed by generic *–DIr* marker ($M = 1.13$, $SD = .13$) more than when it was expressed by inferential use of *-mIş/-(y)mIş* ($M = .64$, $SD = .14$), $p < .05$; but not more than when it is expressed by hearsay use of *–mIş/-(y)mIş* ($M = .80$, $SD = .14$), $p > .05$.; and there was no significant difference between main comparisons in terms of total trials.

Overall, the number of children who chose chose inferential source as the more reliable source is much higher ($N = 77$) than the number of children who chose hearsay source as the more reliable source ($N = 19$), Table 7 presents the means and standard deviations of reliability preference and linguistic marker condition for each age group.

The next hypothes and the corresponding analyses are stated below:

i. In the inferential *-mIş/-(y)mIş* condition, children who attribute higher reliability to the inferential use of *-mIş* will show more generalization than children who attribute higher reliability to the hearsay use of *-mIş*.

ii. In the hearsay *-mIş/-(y)mIş* condition, children who attribute higher reliability to the hearsay use of *-mIş/-(y)mIş* will show more generalization than children who attribute higher reliability to the inferential use of *-mIş*.

Table 7. Means and Standard Deviations of Reliability Preference and Linguistic Marker Condition by Age Group ($N = 96$)

	Ling. Marker	Reliability Pref.	N	Mean	SD
Duration (Seconds-Log10)	*-DIr*	Hearsay Preference	7	.83	.80
		Inference Preference	25	1.43	.45
		Total	32	1.30	.59
	Inferential *-mIş*	Hearsay Preference	6	.40	.63
		Inference Preference	26	.88	.66
		Total	32	.79	.68
	Hearsay *-mIş*	Hearsay Preference	6	.84	.77
		Inference Preference	26	.76	.64
		Total	32	.78	.65
Number of Blickets (Log10)	*-DIr*	Hearsay Preference	7	.29	.28
		Inference Preference	25	.54	.24
		Total	32	.49	.27
	Inferential *-mIş*	Hearsay Preference	6	.22	.42
		Inference Preference	26	.32	.26
		Total	32	.30	.29
	Hearsay *-mIş*	Hearsay Preference	6	.32	.29
		Inference Preference	26	.27	.22
		Total	32	.28	.23
Total Number of Trials (Log10)	*-DIr*	Hearsay Preference	7	.52	.53
		Inference Preference	25	.98	.44
		Total	32	.88	.49
	Inferential *-mIş*	Hearsay Preference	6	.25	.43
		Inference Preference	26	.56	.49
		Total	32	.50	.48
	Hearsay *-mIş*	Hearsay Preference	6	.49	.47
		Inference Preference	26	.48	.45
		Total	32	.48	.45

A two-way MANOVA was planned to examine the effect of reliability preference (2) and linguistic conditions of *-mIş/-(y)mIş* marker (2) on duration, number of blickets and total number of trials as generalization variables. The *-DIr* condition was excluded from the analysis so that only the scores of children who were assessed in the inferential use of *-mIş/-(y)mIş* and hearsay use of *-mIş/-(y)mIş* conditions were compared in terms of reliability preferences and generalization patterns. However, after the *-DIr* condition was excluded, only 12 children, who chose hearsay as the more reliable source remained in the data as opposed to 52 children who chose inferential source as the more reliable source. Given this discrepancy in the number of children in the two categories, neither a MANOVA nor a linear regression analysis was carried out as it would not give statisticaly meaningful results.

3.2.4 The link between Theory of Mind and assessing source reliability

One of the hypotheses was that children who have a higher score on Theory of Mind (ToM) tasks will attribute higher reliability to information expressed with inferential *-mIş/-(y)mIş* than to information expressed with the hearsay *-mIş/-(y)mIş*. A Pearson's correlation analysis between inference reliability score and total ToM score was performed. Results indicated that the predicted relation was not significant, $r(94) = .13$, $p > .05$. A linear regression analysis also demonstrated that children's total ToM scores did not predict their inference reliability scores, $b = .13$, $t(94) = 1.26$, $p > .05$.

Separate Pearson's correlation analyses between inference reliability score and Knowledge Access and Unexpected Contents scores were conducted to see whether there is a significant relation between them, separately. Results showed that

children who have higher scores on the Knowledge Access task assessed inferential source as the more reliable source, $r(94) = .24$, $p < .05$; but there was no significant association between Unexpected Contents task and assessing inferential source as more reliable source, $r(94) = .03$, $p > .05$. The number of children giving correct responses of Knowledge Access task and Unexpected Content task by age are presented separately in Figure 11 and Figure 12, respectively.

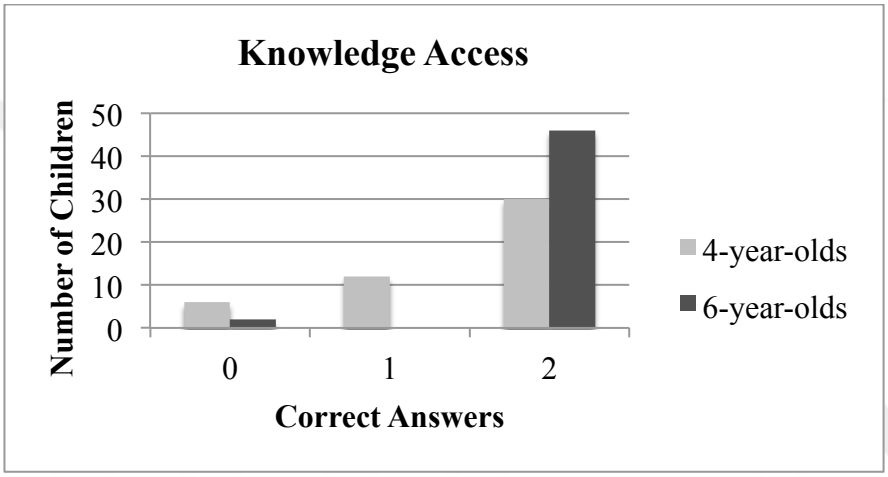

Figure 11. Distribution of children in terms of the number of correct responses they gave on the Knowledge Access task by age

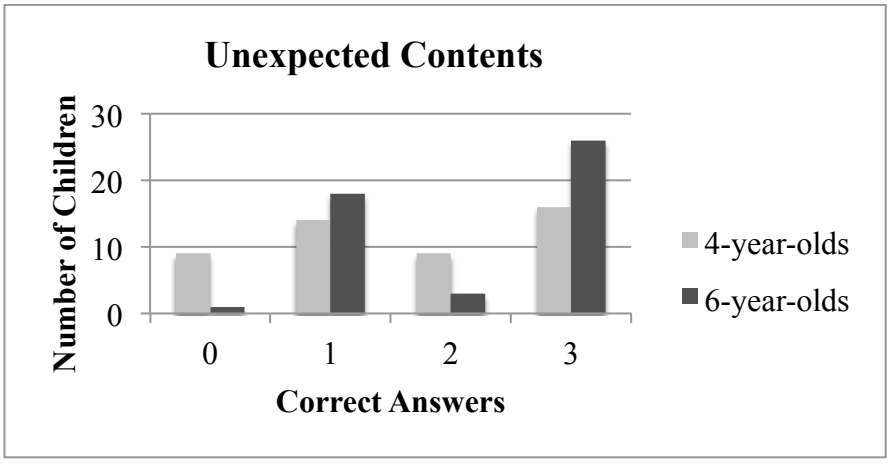

Figure 12. Distribution of children in terms of the number of correct responses they gave on the Unexpected Contents task by age

Another Pearson's correlation analysis was performed to see the relation between age and total ToM score. It demonstrated that older children have higher scores on Theory of Mind tasks than younger children, $r(94) = .36$, $p < .05$. A hierarchical linear regression analyses was performed to see the effect of age, Knowledge Access score and Unexpected Contents score on inference reliability score. According to the linear regression analysis, age predicted the inference reliability score significantly in Step 1, $b = .41$, $t(94) = 4.31$, $p < .001$, with an R^2 of .17. When Knowledge Access score was added as a independent variable in Step 2, age still predicted the inference reliability score significantly, $b = .37$, $t(94) = 3.60$, $p < .01$, with an R^2 of .17, but there was no significant effect of Knowledge Access score, $b = .10$, $t(94) = .38$, $p > .05$. The Unexpected Contents score was added in Step 3, but this score also was not found to be a significant predictor of inference reliability score, $b = .14$, $t(94) = 1.32$, $p > .05$, $b = -.12$, $t(94) = -1.17$, $p > .05$, respectively (See Table 8).

To explore further the significant effect of age, that is, what age might be representing, a hierarchical linear regression was performed with Knowledge Access score and Unexpected Contents score as predictors of inference reliability score. Results indicated that Knowledge Access score predicted the inference reliability score significantly, $b = .24$, $t(94) = 2.42$, $p < .05$, with an R^2 of .06. When Unexpected Contents score added in Step 2, Knowledge Access score still predicted the inference reliability score significantly, $b = .27$, $t(94) = 2.47$, $p < .05$, with an R^2 of .06, but Unexpected Contents score did not predict the inference reliability score, $b = -.07$, $t(94) = -.66$, $p > .05$. The hierarchical regression results are given in Table 9.

Table 8. Hierarchical Linear Regression Analysis Predicting Inference Reliability Score by Age, Knowledge Access Score and Unexpected Contents Score ($N = 96$).

	β	*t*	*Sig.*	R^2	ΔR^2	*F for* ΔR^2
Step 1				.17***	.16***	18.54***
Age	.41	4.31	.000***			
Step 2				.17	.16	.97
Age	.37	3.60	.001***			
Knowledge Access Score	.10	.98	.33			
Step 3				.17	.16	1.37
Age	.38	3.72	.000***			
Knowledge Access Score	.14	1.32	.19			
Unexpected Contents Score	-.12	-.1.17	.25			

The β *weights are the standardized coefficients at each step.* * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 9. Hierarchical Regression Analysis Predicting Inference Reliability Score by Knowledge Access and Unexpected Contents Score ($N = 96$).

	β	*t*	*Sig.*	R^2	ΔR^2	*F for* ΔR^2
Step 1				.06*	.05*	5.84*
Knowledge Access Score	.24	2.42	.02			
Step 2				.06	.04	.43
Knowledge Access Score	.27	2.47	.02			
Unexpected Contents Score	-.07	-.66	.51			

The β *weights are the standardized coefficients at each step.* * $p < .05$, ** $p < .01$.

In order to see whether there is an effect of age and gender on Theory of Mind performance, a two-way ANOVA was performed. Results indicated that 6 years old children had significantly higher scores than 4 years old children, $F(1, 92) = 7.47$, $p < .01$, $\eta p^2 = .08$. There was also a marginally significant trend for gender, girls had higher scores than boys $F(1, 92) = 3.55$, $p = .06$, $\eta p^2 = .04$. However, there was no interaction between age and gender. The means and the standard deviations are presented in Table 10.

Table 10. Means and Standart Deviations of Theory of Mind Scores

Gender	Age Group	N	Mean	SD
Female	4-year-olds	20	3.60	.31
	6-year-olds	29	4.17	.26
Male	4-year-olds	28	2.86	.26
	6-year-olds	19	3.84	.31

3.2.5 The link between TİFALDİ scores and assessing source reliability and Theory of Mind

A Pearson's correlation analysis was conducted to see whether there is a significant relation between a child's TİFALDİ standard score, which indicates general language competence indexed by vocabulary, and inference reliability score, which indicates reliability attribution to inferential sources. No significant relation was found, $r\ (94) = .03$, $p > .05$.

On the other hand, a Pearson correlation analysis indicated that children who have higher TİFALDİ standard scores also have higher total ToM scores, $r(94) = .48$,

$p < .001$, showing a bivariate relationship between ToM and general language competence. When we examined whether performance on each ToM task is related to TİFALDİ standard scores, Pearson's correlations indicated that TİFALDİ standard score is positively associated both with the Knowledge Access score, $r\ (94) = .36$, $p < .001$, and with Unexpected Contents score, $r\ (94) = .43$, $p < .001$. A hierarchical linear regression was run to examine each ToM task's effect on TİFALDİ standard score. Hierarchical regression results are given in Table 11.

Table 11. Hierarchical Regression Analysis Predicting TİFALDİ Standard Score by Knowledge Access Score and Unexpected Contents Score ($N = 96$)

	β	*t*	*Sig.*	R^2	ΔR^2	*F for* ΔR^2
Step 1				.13***	.12***	13.87***
Knowledge Access Score	.36	3.73	.000***			
Step 2				.23**	.21**	11.98**
Knowledge Access Score	.22	2.52	.03*			
Unexpected Contents Score	.34	3.46	.001**			

The β *weights are the standardized coefficients at each step.* * $p < .05$, ** $p < .01$, *** $p < .001$.

Thirteen percent of the variance of TİFALDİ Standard Score is explained by Knowledge Access score, and 23% of the variance is explained by Knowledge Access and Unexpected Contents score, together.

CHAPTER 4

DISCUSSION

The present study aimed to investigate the effect of language on preschool and primary school children's understanding of generic vs. specific knowledge indicated by specific grammticized morphemes and their assessment of source reliability. More specifically, the goal of the study was to explore whether children are able to differentiate between knowledge that is specific to a particular object and knowledge that is generalizable to other instances of the same kind on the basis of linguistic information indicated by epistemic versus evidential morphology and whether the generalization is associated with the assessment of source reliability as indicted by the different functions of the evidential, namely, first hand but partial information versus secondhand information. Four- and 6-year-olds participated in the study and their performance in terms of generalizability and reliability attributions were compared in different tasks. Furthermore, the relation between understanding of linguistic encoding of generalizability and reliability attribution with ToM skills and general language competence indexed by vocabulary were investigated. The discussion is organized around the hypotheses that state the predictions of the study.

First, it was hypothesed that children will generalize the property of an object to other instance of the same kind more, when the property is expressed by use of the generic marker *-DIr* than when it is expressed by the inferential use of *-mIş/-(y)mIş* or the hearsay use of *-mIş/-(y)mIş*. The results supported the hypothesis since children tried the inert blickets for longer duration of time, with more blickets and over more trials to see their magnetic property when the property was expressed by the generic marker *-DIr* than when it was expressed by the inferential or hearsay use

of *-mIş/-(y)mIş*. Overall, this finding is in line with the previous finding of Tamm et al.'s study (2014) which indicates that children generalize more when the linguistic expression uses the generic marker *-DIr* than when the linguistic expression uses the evidential *-mIş/-(y)mIş* marker, regardless of its different functions.

Tamm and her colleagues (2014) demonstrated that both 4-year-old and 6-year-old Turkish children generalize information when the property is communicated linguistically by use of the generic marker *-DIr* without an ostensive demonstration. They also found that children do not generalize when the property is presented with the *-mIş/-(y)mIş* marker and in an unmarked baseline condition. In the current study, 4 years old children generalized the property of the object to other instances of the same kind more when it was expressed by the generic marker *-DIr* than when it was expressed by the inferential or hearsay use of *-mIş/-(y)mIş*. This finding supported the results of Tamm et al.'s study (2014). However, 6 years old children did not show any difference among the generic *-DIr*, the inferential and hearsay *-mIş/-(y)mIş* presentations in terms of generalization since they tested the magnetic property of the object equally regardless of whether the linguistic expression was generic *-DIr,* the inferential or the hearsay use of *-mIş/-(y)mIş*. Six years old children tended to generalize the information to the other instance of a same kind in every linguistic condition. This finding could be explained in terms of several factors. First, there may be a change in the comprehesion of *-mIş/-(y)mIş* marker with age. Previous experimental studies show that in terms of production children do not display full performance before 5 years of age (Aksu-Koç, 1988; Ögel, 2007). Öztürk and Papafragou (2016) found that semantics and pragmatics of evidentiality are not fully developed even until 6 and 7 years of age. Ünal & Papafragou (2016) have further argued that there is an asymmetry between production and comprehension of

evidential markers such that children's comprehension follows their production. They explain this by reference to complex conceptual processes that may be involved in comprehension. Six years old children may be attributing generalizable meaning to evidential *-mIş/-(y)mIş* marker more than 4 years old children do because of the change of their reliability attribution to evidential *-mIş/-(y)mIş* marker with age. Second, they may be hypothesizing about what the experimental situation expects of them and try to meet those expectations by trying more in every condition at this age. Further studies are needed to see what may be the reason.

The second hypothesis predicted that children generalize the property of an object to other instances of the same kind when it is conveyed by the inferential use of *-mIş/-(y)mIş* more than when it is conveyed by the hearsay use of *-mIş/-(y)mIş,* with the assumption that they would treat inferences based on partial direct evidence as more reliable than secondhand information. The results did not support this hypothesis, and demonstrated that children test the magnetic quality of inert blickets for a similar duration of time with similar number of blickets and trials when the property is expressed by the inferential use of *-mIş/-(y)mIş* and the hearsay use of *-mIş/-(y)mIş.* More specifically, 4-year-olds did not generalize in either condition, whereas 6-year-olds generalized in both conditions. The findings, therefore, indicate that children do not differentiate the inferential and hearsay uses of *-mIş/-(y)mIş* in terms of generalizability. However, 6-year-olds attribute more generalizability to the evidential *-mIş/-(y)mIş* marker. This is the first study that tested the the inferential and hearsay use of evidential *-mIş/-(y)mIş* separately in terms of generalization. Therefore, we cannot compare the results with results of other studies in the literature. Further studies are necessary to make a replication.

With the third hypothesis, 6-year-olds were expected to show a higher level of understanding of *-DIr* as a generic marker than 4-year-olds. This hypothesis was not supported since the generalization patterns of 4-year-olds and 6-year-olds did not differ significantly. Both 4- and 6-year-olds generalized in response to the *-DIr* suffix. This finding confirms the results of the Tamm et al. (2014) which also show that *-DIr* has already been interpreted as a generic marker at age of 4. On the other hand, Aksu-Koç & Alıcı (2000) observe that the understanding of *-DIr* as a means of expression of degrees of certainty increases with age. Their study further differs from the present one as it called for metalinguistic judgements, whereas the results of the present study rest on direct actional responses from the children without necessitating reflective thinking.

For source reliability, it was expected that children would attribute higher degree of reliability to evidential information based on inferences from partial observable evidence than to information based on hearsay that indicates secondhand evidence. As expected, it was found that both 4-year-olds and 6-year-olds chose information from partial observable evidence, as more reliable than information based on secondhand evidence. Children's ability to differentiate between the different evidential forms in terms of their implications of reliability was demonstrated by Aydın and Ceci (2009) who showed that 4 years old Turkish children have a capacity to differentiate direct evidence marker *-DI* from indirect evidence marker *-mIş/-(y)mIş* in misinformation contexts. Older Turkish children between 5- to 6 years are fully capable of treating as more reliable a direct evidence than an indirect evidence. Similarly, Fitneva (2008) demonstrated that 6- and 9-years-old Bulgarian children differentiate direct experience marker, hearsay marker and inferential marker in terms of reliability; both groups of children reported direct

perception more reliable than inferential and hearsay expressions. With the results of the present study, we show that Turkish children observe a difference between the two types of evidential expressions, inferential and hearsay, in terms of source reliability even at the age of 4. Both 4- and 6-years-old Turkish children assessed partially observable evidence more reliable than secondhand information. This result may lead us to think that even 4-year-olds can comprehend both inferential and hearsay use of evidential *-mIş/-(y)mIş* marker, because they differentiate these sources in terms of their reliability. This result is against the findings that show no comprehension of evidentials until age of 6 or 7 (Öztürk & Papafragou, 2016). Instead, it may be that children cannot show their comprehension of evidentials because of task demands (Aksu-Koç, 2009). It is clear that further research is needed for a clear picture.

Considering developments in the meaning and implications of evidentials with age, it was hypothesized that 6-year-olds would attribute reliability to the inferential use of *-mIş/-(y)mIş* more than 4-year-olds do. Our findings supported this hypothesis and demonstrated that 6-year-olds attributed reliability to the source who had the information by an inference more than 4-year-olds did. Children assume that the source is more reliable when there is partial observable evidence rather than secondhand knowledge. Previous studies about source reliability showed that even 14-month-olds can differentiate the reliable and unreliable sources (Chow et al., 2008; Poulin-Dubois et al., 2011; Zmyj et al., 2010). The literature indicates that children are capable of differentiating source reliability by using their linguistic ability, for example their knowledge of vocabulary used in labeling objects when they are 3 to 4 years old (Blanco, 2013; Koenig et al., 2004; Lucas et al., 2013). Our findings contribute to this literature by showing the developmental pattern of

assessing source reliability on the bais of more abstract but pervasive grammatical cues such as verb morphology that encode different types of evidence such as inference and hearsay. This is the first study that differentiates the inferential and hearsay use of evidential *-mIş/-(y)mIş* markers in terms of reliability attribution. Therefore, further research is necessary to replicate and generalize these current findings.

In hypothesis six, it was predicted that there would be a positive relationship between reliability attributions and generalizability patterns. However, the analysis that explored the relationship between children's reliability preference and their generalizing performance in the *-DIr*, the inferential and the hearsay *-mIş/-(y)mIş* conditions did not support the hypothesis. After excluding the *-DIr* condition the number of children in the two groups were radically different since majority of the children chose inferential source ($N = 52$) more reliable than hearsay source ($N = 12$), a meaningful analysis could not be conducted.

Another hypothesis that was tested stated that children who have higher scores of Theory of Mind will attribute more reliability to information expressed with inferential *-mIş/-(y)mIş* than to information expressed with the hearsay *mIş/-(y)mIş*. The results did not support this prediction, but separate analyses showed that successful performance on the Knowledge Access task predicts the choice of the more reliable source. Children who displayed an understanding of knowledge access chose inference as the more reliable source whereas successful performance on the Unexpected Content task did not predict reliability attributions. Knowledge Access task measures understanding the relation between knowledge (what is in the box) and its mode of acquisition (seeing) from someone else's different viewpoint whereas Unexpected Content task measures understanding someone else's false belief (Well-

man & Liu, 2004). It is meaningful to find a relation between the ability of assessing others' knowledge access and assessing the reliability of the source since the former is a prerequisite to the latter. Ünlütabak (2012) also indicated a ceiling effect in Knowledge Access scores of 3- to 6-year-old Turkish children reporting higher performance compared to the results of Wellman and Liu's (2004). Our results supported these findings. In this regard Lucas et al.'s study (2013) claimed that 4 years old Turkish children are advantaged both in selective trust and false belief performance thanks to the evidential markers they have in their native language. However they did not directly assess children's performane on evidential tasks. Therefore, further studies are needed to explicate the relations between children's understanding of false belief and their attributions of reliability to evidentials.

Finally, our findings indicated a positive significant relationship between children's vocabulary capacities and Theory of Mind abilities. Analyses showed that both of them have an effect on another. This finding is also in line with previous findings that claim development in language abilities and Theory of Mind abilities affect one another in a reciprocal way (Astington & Baird, 2005).

4.1 Contributions of the present study

This is the first study that differentiates the inferential and the hearsay uses of the *-mIş/-(y)mIş* marker for investigating their effects on generalizability and also on reliability. This study integrated the question of source reliability to the generic knowledge acquisition research. Moreover, it is the first study that investigated whether the types of inferences marked by the inferential use of *-mIş/-(y)mIş* and the hearsay use of *-mIş/-(y)mIş* are applicable only to specific objects or events that have

an episodic nature or whether they are on generalizability context. Finally, this study integrated the relation of Theory of Mind with source reliability.

A new methodology was developed in order to contextually differentiate the inferential use of *-mIş/-(y)mIş* and hearsay use of *-mIş/-(y)mIş* to test generalizability in two different evidential conditions. Moreover, the present study is also novel for investigating and comparing two forms of indirect experience that are partially direct experience (inferential *mIş/-(y)mIş* form) and secondhand information (hearsay *mIş/-(y)mIş* form) in terms of source reliability. A new methodological tool was developed using by Adobe Flash Player program. Animation items were prepared and used to compare source reliability of the two forms of the evidential marker (*-mIş/-(y)mIş*). These animations can be used for assessing source reliability of partial direct evidence and secondhand information in future studies. It is also the first study, which investigates the relationship of source reliability and generic knowledge acquisition. It is informative about Theory of Mind abilities concerning knowledge access and its relation to assesments of reliability of information source.

4.2 Limitations of the study and suggestions for future studies

Although this study contributes to the psycholinguistics literature, there are some limitations. First, we tried new tasks developed for this study to differentiate inferential and hearsay uses of same evidential form (*-mIş/-(y)mIş*). In the Generalizability Task, we wanted to compare the linguistic expressions without an ostensive demonstration, because there is no ostensive demonstration in Tamm et al.'s study (2014). However, we had to show the end result of the magnetic attraction process with a magnetically attached object in the condition of inferential use of *-mIş* marker because we had to give a partial observable evidence for making an

inference. We also had to demonstrate a conversation on the phone in the condition of hearsay use of *-(y)mIş*, because we had to create the context for secondhand information. Our contexts of *-DIr* marker, inferential and hearsay *-mIş* marker in the experiment reflect proper use of these markers in our daily life.

In the present study, only the "generic/certain" meaning of *-DIr* was investigated. However, *-DIr* marker does not only carry a meaning of certainty and / or genericity which indicates a high degree of reliability, but also a noncertain / probabilistic meaning, which indicates lower degrees of reliability (Sansa-Tura, 1986, p. 145; Aksu-Koç & Alıcı, 2000; Aydın & Aksu-Koç, 2015). In future studies, noncertain / probabilistic meaning of *-DIr* can also be operationalized and investigated in terms of generalizability and source reliability. Moreover, only inferential and hearsay functions of *-mIş/-(y)mIş* were compared in terms of reliability. Children's assessment of *-DIr* in terms of the reliability of information it expresses should also be obtained and compared with the reliability attributions to *-mIş/-(y)mIş*. A bivariate relationship was found between children's general language competence and Theory of Mind performance. However, for an interpretation of direction of causality future longitudinal studies are needed since it is not possible to do so with the present concurrent data. Finally, 96 children participated in this study, so there were only 16 participants in each group. Inclusion of more participants in each condition would make the results more powerful. The age range may also be increased to investigate the developmental pattern of comprehension of the evidential marker *-mIş/-(y)mIş* in future studies.

4.3 Conclusion

This study investigated the effect of language on the understanding of generic knowledge and source reliability. It examined why generalization is blocked when the evidential form *-mIş/-(y)mIş* used. Results indicated that 4-year-olds generalized information more when it was conveyed with the generic *-DIr* and did not generalize when it was conveyed with either the inferential or the hearsay uses of *-mIş/-(y)mIş*. However, 6-year-olds generalized the information in all cases. Moreover, children attribute low reliability to the hearsay use of *-mIş/-(y)mIş* and do not generalize information conveyed by a hearsay utterance. Also, children attribute high reliability to inferential use of *-mIş/-(y)mIş* but they also do not generalize information conveyed by an inferential utterance. This finding is taken as evidence indicating that the inferential use of *-mIş/-(y)mIş* conveys information specific to a particular object or event and is therefore not generalizable. However, no significant relationship between reliability and generalizability was found.

APPENDIX A

CHARACTERS AND SCENES OF EVIDENTIARY RELIABILITY TASK

Figure A1. All characters of Evidentiary Reliability task

Figure A2. Hearsay condition scene of *ye*, 'eat' animation

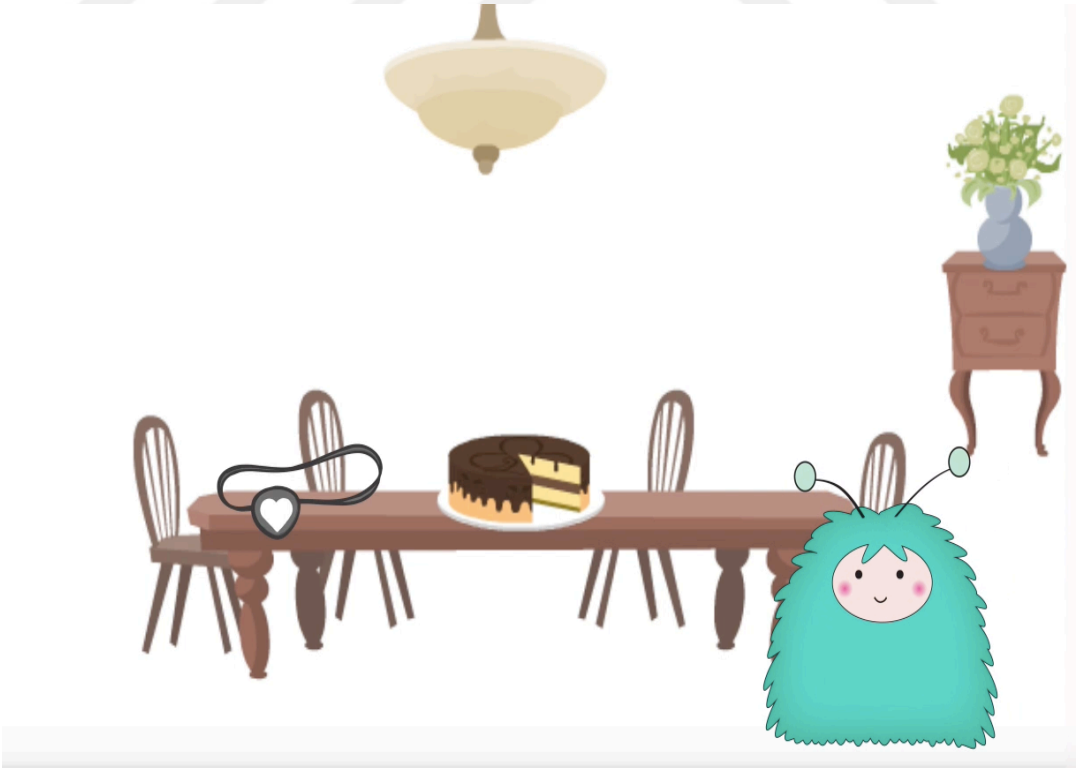

Figure A3. Inferential condition scene of *ye*, 'eat' animation

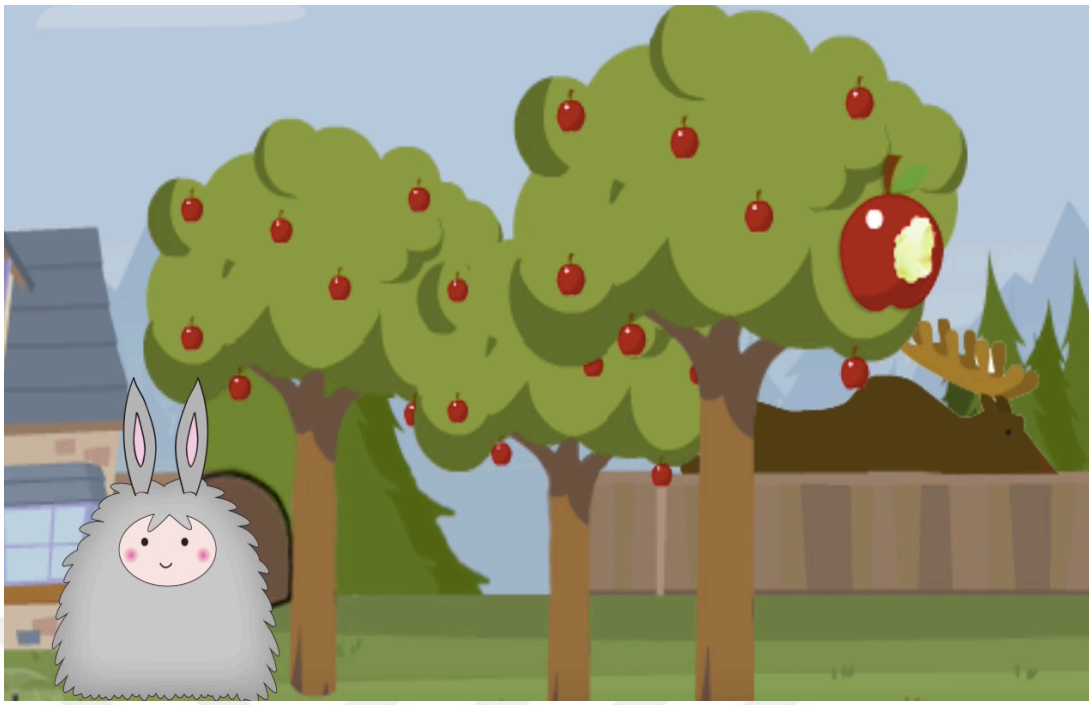

Figure A4. Inferential condition scene of *ɪsɪr*, 'bite' animation

Figure A5. Hearsay condition scene of *ɪsɪr*, 'bite' animation

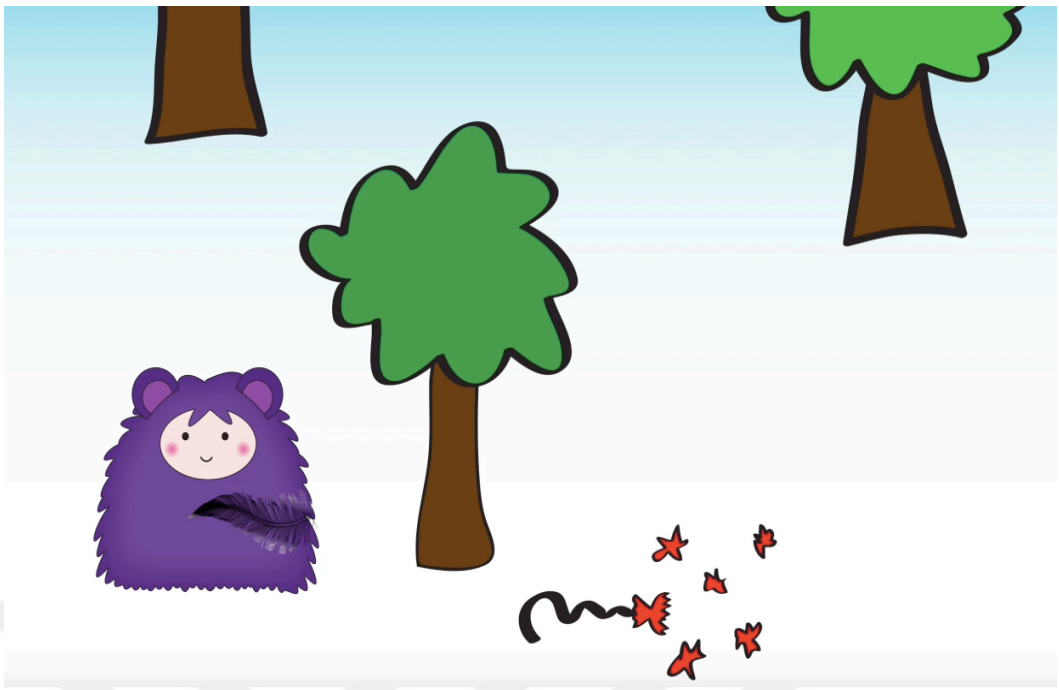

Figure A6. Inferential condition scene of *patlat*, 'pop' animation

Figure A7. Inferential condition scene of *dök*, 'pour' animation

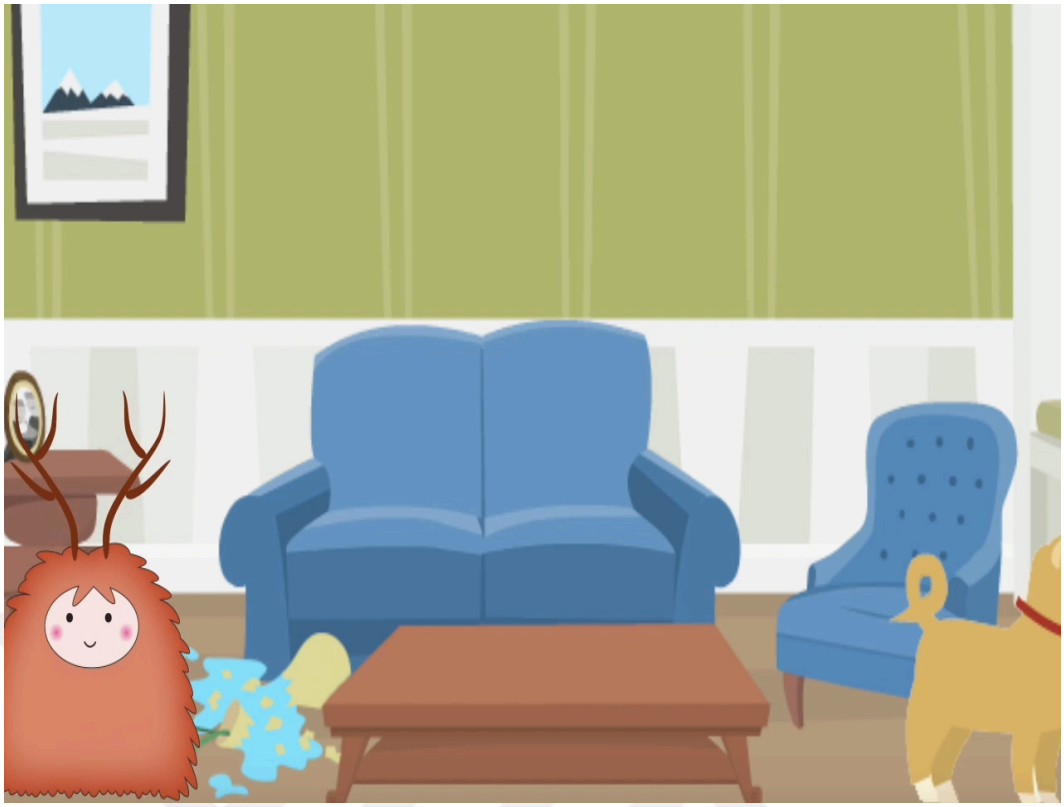

Figure A8. Inferential condition scene of *kır*, 'break' animation

Figure A9. Inferential condition scene of *iç*, 'drink' animation

Figure A10. Hearsay scene of *yırt*, 'tear' animation

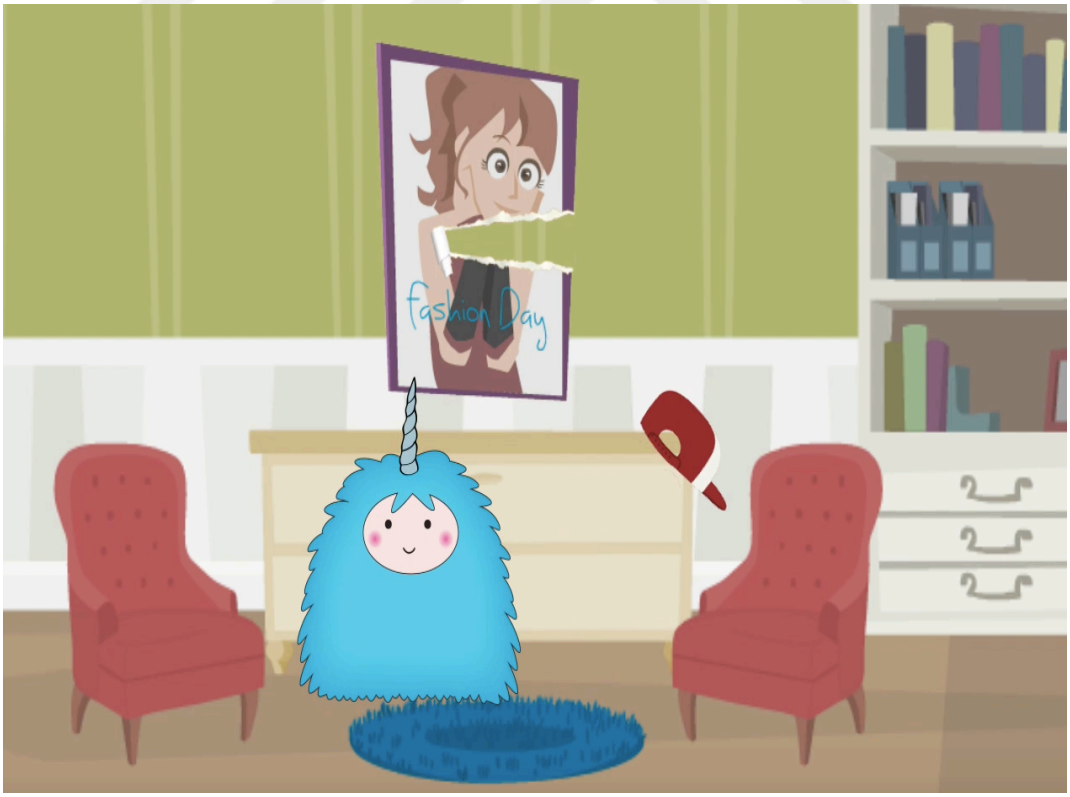

Figure A11. Inferential scene of *yırt*, 'tear' animation

APPENDIX B

KNOWLEDGE ACCESS AND UNEXPECTED CONTENT TASKS (ENGLISH)

I. **Knowledge Access task**

(A closed box, a toy dog ve a doll-boy)

E. Here is a box.

Pre-test question: What do you think is inside of this box?

C.……

E. That is a good guess, let's open and see what it is inside! Look, it is a toy dog inside!

(Toy dog is shown, and then the box is closed again.)

Control Question 1: What is inside of the box?

C. ……

(*The doll-boy enters*)

This boy has not seen the box and has not opened it.

Control Question 2: Did the boy look inside of the box?

C. ……

Target question: Does the boy know what is inside of the box?

C. ……

E. Why does he know / not know?

C. ……

(Keceli Kayisili & Acarlar, 2011)

II. **Unexpected Contents task**

(Candy –Bonibon- box which has pencils inside)

Pre-test question

E. Look! I will show you something. What do you think is inside this box?

C. ……

E. (Experimenter opens the box and shows that there are pencils in the box) What is inside of the box?

C. ……

Target Question 1

E. Yes, there are pencils inside of the box. What did you think was in the box before I opened it?

C. ……

Target Question 2

E. Your friend hasn't seen this box or what is inside of it, yet. I will call her/him shortly. What would s/he think is in the box before I open it?

C. ……

Explanation

E. Why?

C. ……

(Keceli Kayisili & Acarlar, 2011)

APPENDIX C

KNOWLEDGE ACCESS AND UNEXPECTED CONTENT TASKS (TURKISH)

I. Bilgi Edinme (Knowledge Access)

(Kapalı bir kutu, oyuncak köpek ve bebek)

U. Bu bir kutu.

Ön-test sorusu: Sence içinde ne var?

Ç.……

U. Evet güzel bir tahmin. Hadi açalım bakalım içinde ne var! Aa bak, içinde bir köpek var!

(Köpek gösterilir ve tekrar kutuya koyulup kapatılır.)

Kontrol Sorusu 1: Kutuda ne var?

Ç. ……

(*Oyuncak bebek girer*)

Bu çocuk daha önce bu kutuyu hiç görmedi ve hiç açmadı.

Kontrol Sorusu 2: Bu çocuk kutunun içine baktı mı?

Ç. ……

Test Sorusu: Bu çocuk kutunun içinde ne olduğunu biliyor mu?

Ç. ……

U. Neden biliyor/bilmiyor?

Ç. ……

(Keceli Kayisili & Acarlar, 2011)

II. Beklenilmeyen İçerik (Unexpected Contents)

(İçinde kalemler olan şeker –Bonibon- kutusu)

Ön Test Sorusu

U. Bak şimdi sana ne göstereceğim. Sence bu kutunun içinde ne var?

Ç. ……

U. (Kutu açılıp içindekiler çocuğa gösterilir) Aa kutuda ne varmış?

Ç. ……

Test Sorusu 1

U. Evet kutunun içinde kalemler varmış. Peki, ben bu kutuyu açmadan önce sen içinde ne olduğunu sanmıştın?

Ç. ……

Test Sorusu 2

U. Arkadaşın henüz bu kutuyu ve içindekileri görmedi, birazdan arkadaşını çağıracağım. Arkadaşın ben bu kutuyu açmadan önce içinde ne olduğunu sanır?

Ç. ……

Tahminin Açıklanması

U. Neden?

Ç. ……

(Keceli Kayisili & Acarlar, 2011)

APPENDIX D

EXAMPLE PICTURES OF

TURKISH EXPRESSIVE AND RECEPTIVE LANGUAGE (TIFALDI) TEST

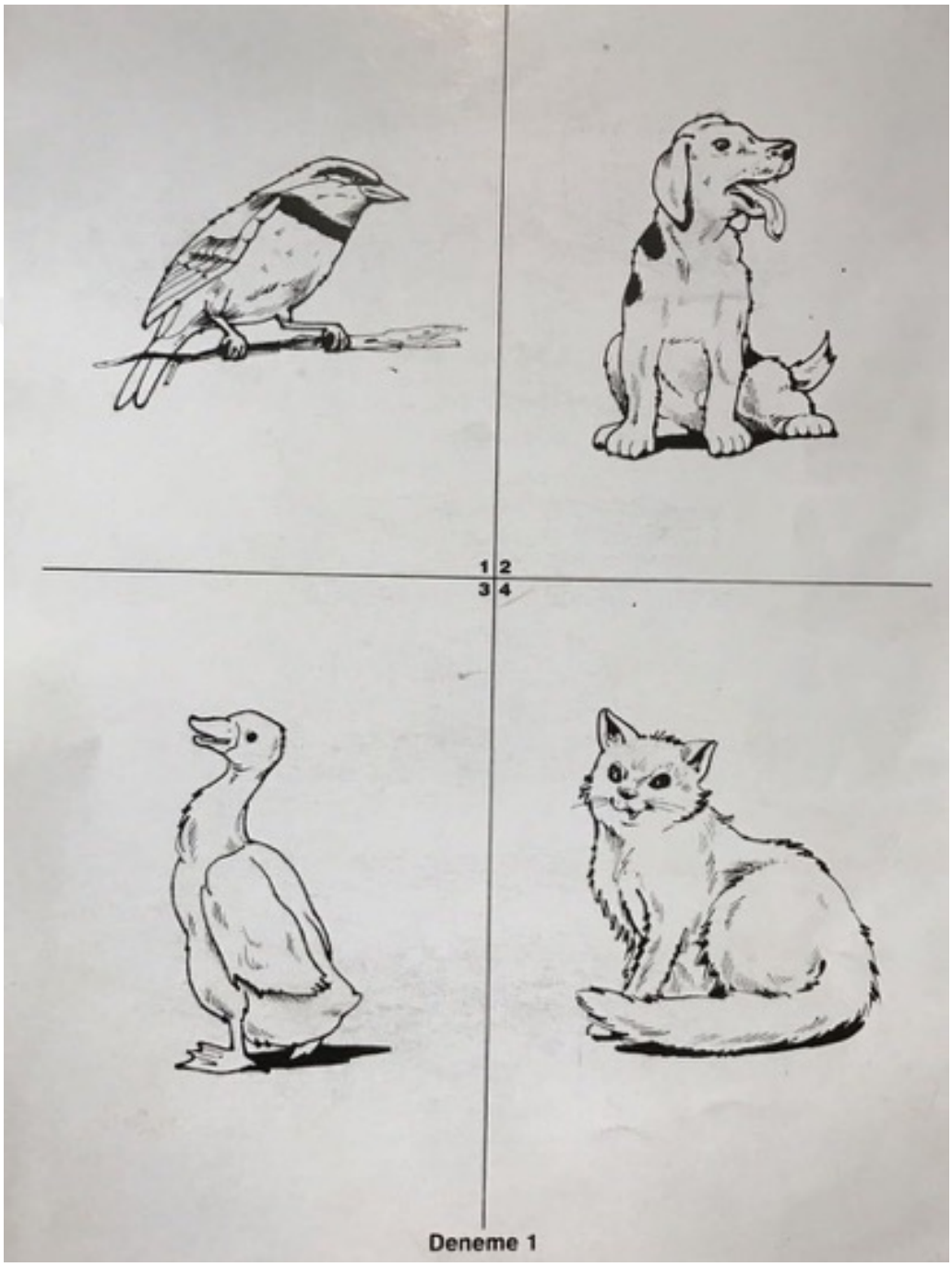

Figure D1. Example pictures of “cat” word to choose correct one

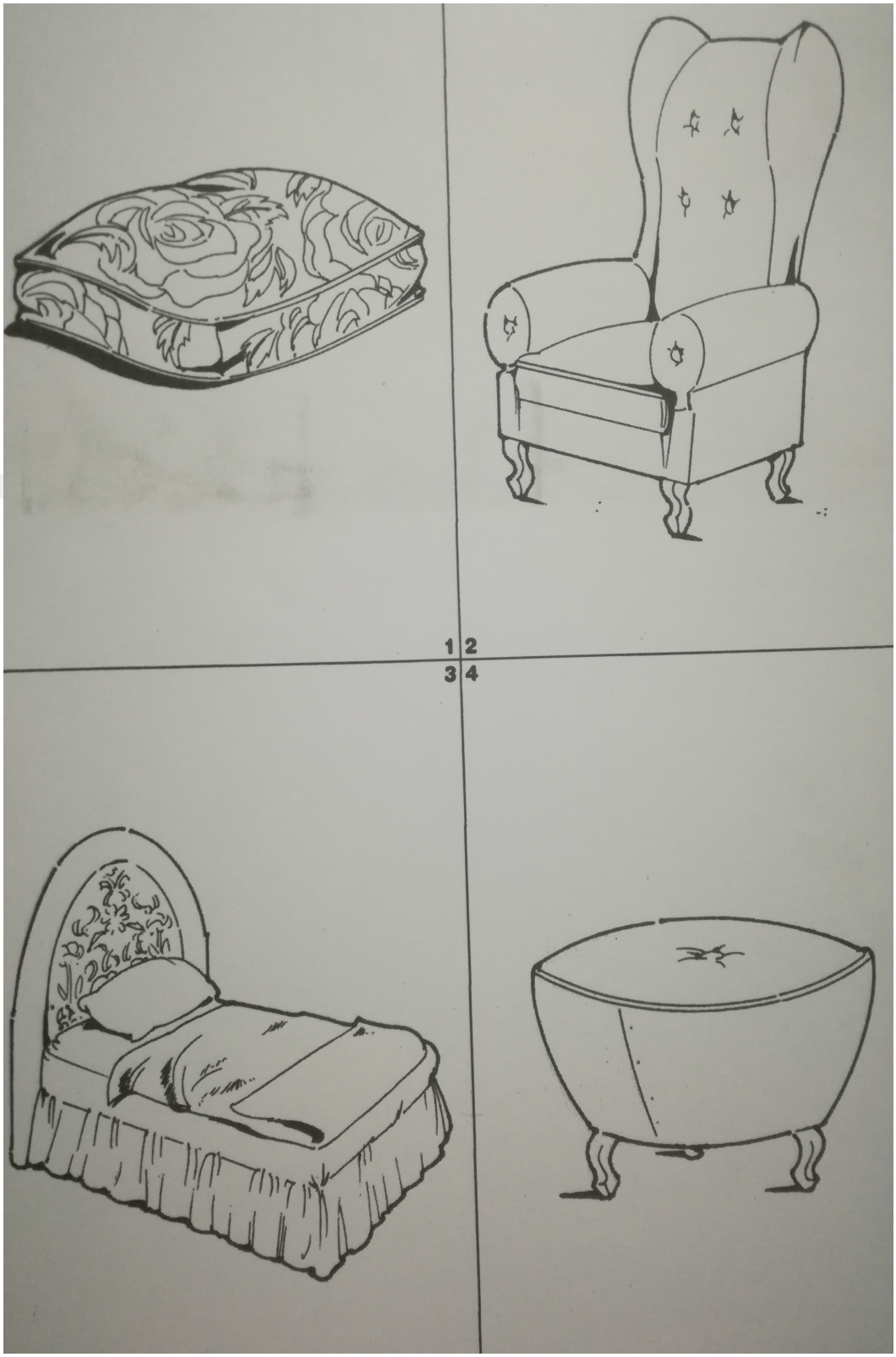

Figure D2. Example pictures of “bed” word to choose correct one

(Kazak Berument & Güven, 2010)

APPENDIX E

DEMOGRAPHIC INFORMATION QUESTIONNAIRE (ENGLISH)

1) Child's name : Sex :	F / M
2) Birthday (day/month/year) :	__ / __ / ____
3) Siblings :	___ No ___ Older than himself / herself ___ Younger than himself / herself
4) The highest education level that the mother completed :	___ Primary ____ Secondary ___High School ____ College __ Master Degree ____PhD.
5) Does the mother work? If the answer is yes, what is the occupation?	__ No __ Yes (…….)
6) The highest education level that the father completed :	___ Primary ____ Secondary ___High School ____ College __ Master Degree ____PhD.
7) Does the father work? If the answer is yes, what is the occupation?	___No ___Yes (……..…)
6) If there is any baby-sitter, please define:	___ No ___ Relatives (Grandmother-aunt) ___ Turkish speaking nanny ___ Foreign speaking nanny
7) Is there any other language speaking at your home? If there is, please define :	___ No ___ English ___ French __German ___ Arabic ___Italian __ Other

8) How many hours is your child exposed to this language?	___ None ___ 0-1 hours __ 1-2 hours ___2-3 hours ___ 3-4 hours __+ 4 hours
9) Do you read a book to your child? If you do, please define the frequency and the language of it:	___ No ___ Everyday ___A few times in a week ___ Once in a week ___A few times in a month ___Once in a month ___ Turkish ___ English ___ Others
10) Your mothly –total- income : (If you do not chose one of the options, please leave it empty)	___ Less than 4000 TL ___ 4000 TL - 6000 TL ___ 6000 TL - 8000 TL ___ 8000 TL- 10.000 TL ___ 10.000 TL - 15.000 TL ___ More than 15.000 TL

APPENDIX F

DEMOGRAPHIC INFORMATION QUESTIONNAIRE (TURKISH)

1)Çocuğunuzun Adı : Cinsiyeti :	K / E
2) Doğum tarihi (gün/ay/yıl) :	__ / __ / ____
3) Kardeşleri :	___ Yok ___ Kendinden büyük kardeş sayısı ___ Kendinden küçük kardeş sayısı
4) Annenin tamamladığı en yüksek eğitim derecesi :	___ İlkokul ____ Ortaokul ___ Lise ____ Üniversite ___ Yüksek lisans ___Doktora
5) Anne çalışıyor mu? Yanıtınız evet ise mesleği nedir?	__ Hayır __ Evet (…….)
6) Babanın tamamladığı en yüksek eğitim derecesi :	___ İlkokul ____ Ortaokul ___ Lise ____ Üniversite ___ Yüksek lisans ___Doktora
7) Baba çalışıyor mu? Yanıtınız evet ise mesleği nedir?	___Hayır ___Evet (……..…)
6) Çocuğunuza sizden başka bakan birisi varsa, lütfen tanımlayınız :	___ Yok ___ Akraba (Büyükanne-teyze) ___ Türkçe konuşan bakıcı ___ Yabancı dilde konuşan bakıcı
7) Evinizde Tükçe'den başka dil kullanılıyor mu? Kullanılıyorsa, lütfen hangi dil olduğunu belirtiniz :	___ Hayır ___ İngilizce ___ Fransızca ___Almanca ___ Arapça ___İtalyanca ____ Diğer

8) Çocuğunuz bu yabancı dile günde kaç saat maruz kalıyor?	___ Hiç ___ 0-1 saat __ 1-2 saat ___2-3 saat ___ 3-4 saat __+ 4 saat
9) Çocuğunuza kitap okuyor musunuz? Okuyorsanız, hangi sıklıkta ve hangi dilde olduğunu belirtiniz.	___ Hayır ___ Her gün ___Haftada birkaç ke ___ Haftada 1 __Ayda birkaç kez ____ Ayda 1 kez ___ Türkçe ___ İngilizce ___ Diğer
10) Aylık toplam geliriniz : (Doldurmak istemediğiniz takdirde boş bırakabililirsiniz)	___ 4000 TL'den az ___ 4000 TL - 6000 TL ___ 6000 TL - 8000 TL ___ 8000 TL- 10.000 TL ___ 10.000 TL - 15.000 TL ___ 15.000 TL'den fazla

REFERENCES

Aikhenvald, A. Y. (2004). *Evidentiality.* New York: Oxford University Press.

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). The role of input vs. universal predispositions in the emergence of tense-aspect morphology: Evidence from Turkish. *First Language*, *18*(54), 255-280.

Aksu-Koç, A. (2000). Some aspects of the acquisition of evidentials. In Lars Johanson & Bo Utas (Eds.), *Evidentials: Turkic, Iranian and neighbouring languages*, (pp. 15-28). Berlin: Mouton de Gruyter.

Aksu-Koç, A. (2009). Evidentials: An interface between linguistic and conceptual development. In D.I Slobin, J. Guo, E. Lieven, S. Ervin-Tripp, N. Budwig, & Ş. Özçalışkan (Eds.), *Crosslinguistic approaches to the psychology of language: research in the tradition of Dan Isaac Slobin* (531–541). Mahwah NJ: Lawrence Erlbaum Associates.

Aksu-Koç, A. (2016). The interface of evidentials and epistemics in Turkish. In M.Güven, D. Akar, B. Öztürk, & M. Kelepir (Eds.), *Exploring the Turkish linguistic landscape*. *Essays in honour of Eser Erguvanlı-Taylan*. Amsterdam: John Benjamins.

Aksu-Koç, A. & Alıcı, D.M. (2000). Understanding sources of beliefs and marking of uncertainty: The child's theory of evidentiality. In E.V. Clark (Ed.), *The Proceedings of the Thirtieth Child Language Research Forum,* (123–130). Stanford CA: Stanford University Press.

Aksu-Koç, A., Aydın, Ç., Avcı, G., Sefer, N., & Yaşa, Y. (2005). The relation between mental verbs and theory of mind performance: Evidence from Turkish children. Paper presented at the *X. International Association for the Study of Child Language,* 2005, Berlin.

Aksu-Koç, A., Ogel-Balaban, H., & Alp, I. E. (2009). Evidentials and source knowledge in Turkish. *New Directions for Child and Adolescent Development*, *2009* (125), 13-28.

Aksu-Koç, A. & Slobin, D. I. (1985).The acquisition of Turkish. In Dan. I. Slobin (Ed.), *The crosslinguistic study of language acquisition, Vol. 1: The data* (pp. 839-878). Hillsdale, NJ: Lawrence Erlbaum Associates.

Aksu-Koç, A., & Slobin, D. I. (1986). A psychological account of the development and use of evidentials in Turkish. In W. Chafe, & J. Nichols, (Eds.) *Evidentiality: The linguistic coding of epistemology*, (159-167.) Norwood, NJ: Ablex.

Astington, J. W., & Baird, J. A. (Eds.). (2005). *Why language matters for theory of mind*. Oxford: Oxford University Press.

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.

Aydın, Ç., & Aksu-Koç, A. (2015). *Children's knowledge acquisition: Comprehension of the epistemic function of the marker -DIr in Turkish*. Poster presented at the Budapest CEU Conference on Cognitive development (BCCCD 2015) January 8–10.

Aydın, Ç., & Ceci, S. J. (2009). Evidentiality and suggestibility: A new research venue. *New Directions for Child and Adolescent Development*, *2009* (125), 79-93.

Blanco, C. P. (2013). *Reliability versus affiliation: selective trust in accented speakers* (Unpublished master's thesis). Texas: University of Texas.

Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration. *Child Development*, *83*(4), 1416-1428.

Butler, L. P., & Markman, E. M. (2014). Preschoolers use pedagogical cues to guide radical reorganization of category knowledge. *Cognition*, *130*(1), 116-127.

Buttelmann, D., Carpenter, M., Call, J., & Tomasello, M. (2007). Enculturated chimpanzees imitate rationally. *Developmental Science, 10*, F31–F38.

Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development*, *84*(2), 422-428.

Carlson, S. M., Koenig, M. A., & Harms, M. B. (2013). Theory of mind. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(4), 391-402.

Chow, V., Poulin‐Dubois, D., & Lewis, J. (2008). To see or not to see: Infants prefer to follow the gaze of a reliable looker. *Developmental Science*, *11*(5), 761-770.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, *13*(4), 148-153.

Fitneva, S. A. (2008). The role of evidentiality in Bulgarian children's reliability judgments. *Journal of Child Language*, *35*(04), 845-868.

Fitneva, S. A. (2009). Evidentiality and trust: The effect of informational goals. *New Directions for Child and Adolescent Development*, *2009*(125), 49-62.

Flavell, J. H. (1999). Cognitive development: Children's knowledge about the mind. *Annual Review of Psychology*, *50*(1), 21-45.

Hogrefe, G. J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, 567-582.

Kazak-Berument, S., & Güven, A. (2010). *Türkçe İfade Edici ve Alıcı Dil Testi (TİFALDİ)*. Istanbul: Türk Psikologlar Derneği.

Keceli Kaysili, B., & Acarlar, F. (2011). The development of Theory of Mind according to false belief performance of children ages 3 to 5. *Educational Sciences: Theory and Practice*, *11*(4), 1821-1826.

Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*, *15*(10), 694-698.

Lucas, A. J., Lewis, C., Pala, F. C., Wong, K., & Berridge, D. (2013). Social-cognitive processes in preschoolers' selective trust: Three cultures compared. *Developmental Psychology*, *49*(3), 579.

Matsui, T., Yamamoto, T., & McCagg, P. (2006). On the role of language in children's early understanding of others as epistemic beings. *Cognitive Development*, *21*(2), 158-173.

Meltzoff, A. N. (1988). Infant imitation after a 1-week delay: Long-term memory for novel acts and multiple stimuli. *Developmental Psychology, 24*, 470–476.

Ögel, H. (2007*). Developments in source monitoring and linguistic encoding of source* (Unpublished master's thesis). Bogazici University, Istanbul, Turkey.

Öztürk, O., & Papafragou, A. (2007). Children's acquisition of evidentiality. In *Proceedings from the 31st annual Boston University conference on language development*. Boston, MA: Boston University.

Ozturk, O., & Papafragou, A. (2016). The acquisition of evidentiality and source monitoring. *Language Learning and Development*, *12*(2), 199-230.

Palmer, F. R. (2001). *Mood and modality*. Cambridge: Cambridge University Press.

Papafragou, A., Li, P., Choi, Y., & Han, C. H. (2007). Evidentiality in language and cognition. *Cognition*, *103*(2), 253-299.

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

Perner, J., Leekam, S. R., & Wimmer, H. (1987). Three-year-olds' difficulty with false belief: The case for a conceptual deficit. *British Journal of Developmental Psychology*, *5*(2), 125-137.

Pillow, B. H. (1989). Early understanding of perception as a source of knowledge. *Journal of Experimental Child Psychology*, *47*(1), 116-129.

Poulin-Dubois, D., Brooker, I., & Polonia, A. (2011). Infants prefer to imitate a reliable person. *Infant Behavior and Development*, *34*(2), 303-309.

Prasada, S. (2000). Acquiring generic knowledge. *Trends in Cognitive Sciences*, *4*(2), 66-72.

Pratt, C., & Bryant, P. (1990). Young children understand that looking leads to knowing (so long as they are looking into a single barrel). *Child Development*, *61*(4), 973-982.

Sansa-Tura, S. (1986). DIr in modern Turkish. In A. Aksu-Koç, & E.E. Taylan,, (eds), *Proceedings of the Turkish Linguistics Conference, August 9–10, 1984* (145–158). Istanbul: Boğaziçi University Publications.

Tamm, A., Çaglar, L. R., Aksu Koç, A., & Csibra, G. (2014). *Generic grammar makes Turkish preschoolers generalize statements and tolerate exceptions.* Poster presented at the Budapest CEU Conference on Cognitive development (BCCCD 2014) January 9–11.

Terziyan, T. (2013). *Acquisition of modality in Turkish* (Unpublished master's thesis). Boğaziçi University, Istanbul, Turkey.

Uzundag, B., Tasci, S., Küntay, A. C., & Aksu-Koç, A. (2016). Functions of evidentials in Turkish child and child-directed speech in early child-caregiver interactions. In *Proceedings of the 40th Annual Boston University Conference on Language Development*.

Ünal, E., & Papafragou, A. (2016). Production–comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language*, *89*, 179-199.

Ünlütabak, B. (2012). *Developmental relations between Turkish preschool children's theory of mind skills and their ability to track character references in narative discources (Unpublished master's thesis).* Boğaziçi University, Istanbul, Turkey.

Wellman, H. M., & Estes, D. (1986). Early understanding of mental entities: A reexamination of childhood realism. *Child Development*, 910-923.

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child Development*, *75*(2), 523-541.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103-128.

Whorf, B. L. (1956). *Language, Thought and Reality*. Boston, MA: MIT Press.

Zmyj, N., Buttelmann, D., Carpenter, M., & Daum, M. M. (2010). The reliability of a model influences 14-month-olds' imitation. *Journal of Experimental Child Psychology*, *106*(4), 208-220.