



RESEARCH ARTICLE

Children’s Understanding of Source Reliability and Knowledge Generalizability from Grammatical Cues: Evidence from Turkish

Merve Ataman-Devrim¹ , Gaye Soley²  and Ayhan Aksu-Koç¹

¹Department of Psychology, Boğaziçi University, Istanbul, Türkiye and ²Department of Cognition, Development and Educational Psychology, Institute of Neuroscience, University of Barcelona, Barcelona, Spain

Corresponding author: Gaye Soley; Email: gayesoley@ub.edu

(Received 04 March 2025; revised 02 October 2025; accepted 30 October 2025)

Abstract

This study investigates Turkish-speaking children’s reliability attributions to linguistic indicators of evidential source and whether source reliability has an effect on knowledge generalizability. Ninety-six four- and six-year-olds were first asked to perform a reliability judgement task where informants used the indirect evidential marker *-mİş* in the contexts of inference and hearsay. Next, they were randomly assigned to three groups and introduced a novel object “*blicket*” declared to be magnetic, using inference, hearsay, and generic statements, and their generalization behaviours were measured. Results showed that both four- and six-year-olds attributed higher reliability to inference compared to hearsay as evidential source, and six-year-olds did so more than four-year-olds. Four-year-olds generalized more in response to generic statements than inferential or hearsay statements, whereas six-year-olds generalized similarly in all conditions. Although children attributed more reliability to inference than hearsay, they did not generalize inferential statements more than hearsay statements.

Keywords: source reliability; evidentiality; hearsay; inference; generic knowledge

Özet

Bu çalışma, Türkçe konuşan çocukların kanıtsal kaynağın dilsel göstergelerine yönelik güvenilirlik atıflarını ve kaynak güvenilirliğinin bilgi genellenabilirliği üzerinde bir etkisi olup olmadığını araştırmaktadır. Doksan altı dört ve altı yaşındaki çocuktan ilk olarak, dolaylı kanıt eki *-mİş*’in çıkarım ve sözel-bildirimden-edinilen-bilgi bağlamlarında kullanımını güvenilirlik açısından değerlendirmeleri istenmiştir. Daha sonra çocuklar rastgele yöntemle üç gruba ayrılmış ve uyaran olarak çıkarım *-mİş*, sözel-bildirimden-edinilen-bilgi *-mİş* ve genelleme *-DIr* ifadelerinin kullanıldığı bu üç koşulda kendilerine “bilikit” adlı yeni bir oyuncak ‘mıknatıslı’ olarak tanıtılmış ve genelleme davranışları ölçülmüştür. Sonuçlar, hem dört hem de altı yaşındaki çocukların kanıtsal kaynak olarak sözel-bildirimden-

edinilen-bilgiye kıyasla çıkarıma daha yüksek güvenilirlik atfettiğini ve altı yaşındaki çocukların çıkarımı dört yaşındakilerden daha güvenilir bulduğunu göstermiştir. Dört yaşındaki çocuklar, çıkarımsal veya sözel-bildirimden-edinilen-bilgi ifadelerine kıyasla genel ifadelere daha fazla genelleme yaparken, altı yaşındaki çocuklar tüm koşullarda benzer düzeyde yüksek genelleme yapmışlardır. Çocuklar çıkarıma, sözel-bildirimden-edinilen-bilgiye göre daha fazla güvenilirlik atfettiler de, çıkarımsal ifadeleri sözel-bildirimden-edinilen-bilgiye kıyasla daha fazla genellememişlerdir.

Anahtar Kelimeler: kaynak güvenilirliği; kanıtsallık; sözel-bildirimden-edinilen-bilgi; çıkarım; genel bilgi

1. Introduction

Critically evaluating knowledge sources and available evidence is an increasingly important ability, as information is now more accessible than ever. This ability is particularly crucial for young children, who, compared to adults, have much less experience and frequently have to depend on others' testimony for knowledge acquisition. While children habitually rely on the testimony of others for learning about the world, past research suggests that they do not indiscriminately endorse the information they receive (Harris et al., 2018; Koenig & Sabbagh, 2013). Rather, from an early age, children are sensitive to different communicative acts and linguistic cues that allow them to make inferences about whether the information they are given is accurate or reliable, and whether it can be generalized to other situations (Csibra & Gergely, 2009; Prasada, 2000).

Linguistic cues that have implications for the reliability of what is communicated are epistemic markers that express how certain the speaker is about the factual status of the information conveyed and evidential markers that indicate the source for its factual status (Palmer, 2001). Languages differ in whether they express these notional categories obligatorily with grammatical means such as verb affixes and particles, or optionally with lexical means such as verbs and adverbs. Children acquiring either type of language learn these epistemic and evidential cues and are able to assess the reliability of statements so marked at an early age. They are sensitive to whether the information conveyed is based on the speaker's direct experience or has been accessed indirectly through someone else's report, that is, hearsay, and they trust the former as a source (e.g., Aydin, 2011; Matsui & Miura, 2009; Ozturk & Papafragou, 2016; Ünal & Papafragou, 2016). As children learn more from more reliable sources relative to less reliable ones, these linguistic cues may influence their knowledge generalizations in their everyday lives. However, we do not have any knowledge regarding children's reliability assessment of the more nuanced distinctions between types of indirect evidence such as inference from observed results of a nonwitnessed process versus hearsay, and their tendency to generalize knowledge based on these sources. Turkish, which grammaticalizes inference versus hearsay as finer distinctions under evidentiality, and genericity versus possibility under epistemic modality, provides an opportunity to ask whether children find knowledge from these indirect sources equally trustworthy and generalizable. In the present study, we explore Turkish-speaking children's reliability attributions to types of indirect evidential source and whether reliability has a differential bearing on the generalizability of information so obtained.

In what follows, we first present an overview of the literature on children's understanding of source reliability and knowledge generalizability. Next, we briefly describe the Turkish evidential and epistemic markers under study and summarize the relevant research on their acquisition in Turkish.

2. Children's understanding of source reliability and knowledge generalizability

The effectiveness of human communication rests on the reliability of the messages conveyed, which, in turn, depends on the credibility of their source. Children as young as 14 months are able to differentiate reliable and unreliable sources of information based on a variety of cues, such as the informant's emotional expression (Poulin-Dubois et al., 2011), facial and linguistic expression (Buttelmann & Zmyj, 2020), use of the same language (Buttelmann et al., 2013), and previous reliability (Zmyj et al., 2010). Older children rely on their first-hand observations compared to the testimony of others; most four-year-olds, but few three-year-olds trust the reliable informant unless their testimony conflicts with what they themselves have seen (Clement et al., 2004). They show selective trust in the reliable informant with previous accuracy (Koenig et al., 2004) and use information from them to make predictions and ask questions (Koenig & Harris, 2005a, 2005b; Scofield & Behrend, 2008). Four-year-olds can also track both semantic (object names) and morphological (noun plurals) accuracy of the informants and trust the previously correct one (Corriveau et al., 2010). Three- and four-year-olds credit more knowledge to speakers who can verify their claims relative to speakers using non-verifiable statements (Butler et al., 2020; Koenig et al., 2015) and trust the testimony of those who have visual evidence over those who have not (Robinson et al., 1999). By age six, children find speakers who can verify their claims more trustworthy regardless of the individual's previous reliability (Butler et al., 2020), and they endorse verbal report claims supported by primary (first-hand) than secondary sources (Aboody et al., 2022).

Children are also sensitive to abstract linguistic cues that signal degrees of speaker certainty and types of information source, both of which have implications for the reliability of the knowledge conveyed. Like most Indo-European languages, English marks these distinctions lexically, using modal (e.g., *may*, *must*) and mental verbs (e.g., *think*, *know*) for epistemic qualifications and mainly adverbs (e.g., *evidently*, *reportedly*) for evidential distinctions. English-speaking three-to-four-year-olds use *think* to mean uncertainty (Diessel & Tomasello, 2001; Shatz et al., 1983), four-year-olds understand the contrast between an unmarked statement and a statement with *may* expressing possibility (Hirst & Weil, 1982), as well as the contrast between *know* versus *think* (Moore et al., 1989), and between *must* and *might* (Leahy & Zalnieriunas, 2021), whereas three-year-olds do not. Five-year-olds understand the contrastive force of modal verbs and evaluate the trustworthiness of speakers on that basis, and nine-year-olds show adult-like performance (Noveck et al., 1996).

In languages such as Korean, Japanese, Bulgarian, and Turkish, on the other hand, epistemic and evidential distinctions are grammaticalized; specific verbal affixes and particles are used to mark direct experience, inference from observed results of non-witnessed processes, hearsay, and assumptions as types of source for the asserted information. Children learn the distinctions these forms make early on (Matsui & Fitneva, 2009). For example, Japanese three-year-olds have a fairly good understanding of certainty expressed by sentence final particles (Matsui et al., 2006), four-year-olds choose the more reliable speaker, but only five-to-seven-year-olds can justify their choices (Matsui & Miura, 2009). Korean children similarly use the sentence ending particles for direct evidence/certainty by age two, for indirect evidence/hearsay by age three, and those for inference by age four (Choi, 2021), whereas they perform at chance level in distinguishing direct evidence utterances from those of hearsay on experimental tasks (Papafragou et al., 2007). Turkish 4-year-olds find statements based on direct perception

to be more reliable than indirect information inferred from the observed results of a nonwitnessed process or information obtained through hearsay (Aydin, 2011; Ozturk & Papafragou, 2016). Bulgarian-speaking six- and nine-year-olds also prefer first-hand information expressed with direct perception markers more than information expressed with hearsay and inferential markers, but only nine-year-olds provide reliability judgments (Fitneva, 2008).

Children not only have to distinguish between reliable and unreliable information, but also understand when the information they have learned is generalizable to other circumstances. To make inferences about the generalizability of knowledge acquired from the testimony of others, children, again, make use of communicative and linguistic cues (Butler & Tomasello, 2016; Chambers et al., 2008; Koenig et al., 2015). Communicative acts using ostensive cues in pedagogical contexts are particularly effective for learning about generalizable information (Butler et al., 2015; Butler & Markman, 2012; Csibra & Gergely, 2009). English-learning two-year-olds respond to linguistic cues such as generic and nongeneric noun phrases for distinguishing between generic and specific information and three-year-olds use both linguistic form and pragmatic context to make generic inferences (Butler & Tomasello, 2016; Gelman & Raman, 2003). Four-year-olds generalize knowledge when generic rather than non-generic sentences are used even in the presence of conflicting evidence (Chambers et al., 2008). Three- and four-year-olds are less tolerant to informant errors concerning semantic information that is generalizable (e.g., object labels) compared to errors concerning episodic information which is event-specific (e.g., object locations) (Stephens & Koenig, 2015). Four-to-seven-year-olds believe that facts presented with a generic linguistic format (e.g., “Hedgehogs eat hexapods.”) are more widely known compared to facts presented with a non-generic linguistic format (e.g., “Last night, this hedgehog ate a hexapod.”) (Cimpian & Scott, 2012, p.422). By age seven, children credit more knowledge to speakers using generic statements (about “Pangolins”) over speakers using nongeneric statements (about “This pangolin”) regardless of the verifiability of the statement (Koenig et al., 2015).

To sum up, by age four, children treat direct evidence as more trustworthy than indirect evidence and are sensitive to linguistic indicators of generalizable knowledge. In the current study, we explore the relations between Turkish-speaking children’s understanding of knowledge reliability and generalizability. Before we introduce our study, we give a brief description of evidentiality and generic marking in Turkish and review prior work relevant to the reliability of Turkish evidentiality markers.

3. Indicators of information source and knowledge generalizability in Turkish

In Turkish, the verb root is followed by a string of affixes that indicate voice, negation, tense-aspect-mood, and person-number. Evidentiality and epistemicity are expressed through the multifunctional tense-aspect-mood suffixes attached to finite verbs and non-verbal predicates. The particular choice and order of these suffixes render different meanings compositionally (Taylan, 2001). When talking about the past, Turkish-speakers are required to make a choice between two inflections, *-DI* and *-mİş*, that contrast in terms of mode of access to information, that is, evidentiality (Aksu-Koç, 1988; Johanson, 2000; Slobin & Aksu-Koç, 1982).¹ The past tense suffix *-DI* is neutral as it expresses first-hand

¹ Affixes alternate according to the rules of vowel harmony, consonant assimilation, and other regular morphophonological processes. The two suffixes *-DI* and *-mİş* are realized as *-di*, *-dı*, *du*, *dü*, *ti*, *-tı*, *tu*, *tü*,

experience as well as knowledge accepted as certain and factual (example 1). It contrasts with the evidential *-mİş* which indicates information obtained indirectly (example 2) either through inference from observed results of a nonwitnessed process (a) or through the verbal report of someone else, that is, hearsay (b). *-DI* gains its evidential force due to the obligatory choice that speakers make between the two forms.

- (1) *Kedi süt-ü dök-tü.*
cat milk-ACC spill-PAST
“The cat spilled the milk (I saw it).”
- (2) *Kedi süt-ü dök-müş.*
cat milk-ACC spill-PAST.EVID
“The cat spilled the milk, evidently.”
(a) I saw paw prints next to the spilled milk (I inferred it from evidence).
(b) I was told that it spilled the milk (I heard it from someone).

To the extent that the use of evidential markers for information source also indicates “speaker stance” (Aksu-Koç, 2016; Johanson, 2000), evidentiality closely interfaces with epistemic modality in Turkish. Of interest here is the epistemic marker *-DİR*, which denotes categorical, certain, hence generalizable knowledge taken to be factual (example 3). As multifunctional suffixes, *-DİR* and compositionally *-mİş-DİR* are also used to express assumptions about probable/possible states of affairs (example 4) based on the speaker’s reasoning from general knowledge and habitualities in the world in the absence of evidence (Aksu-Koç & Alıcı, 2000; Palmer, 2001; Tura, 1986).

- (3) *Kedi evcil bir hayvan-dır.*
cat domestic one animal-GEN
“Cat is a domestic animal.” (generic information)
- (4) *Kedi süt-ü dök-müş-tür.*
cat milk-ACC spill-PERF-POSS
“The cat must have spilled the milk (I assume that/ as it habitually does).”

It is important to note that the interpretation of these multifunctional suffixes is a function of the linguistic and situational context. Whether a statement with *-mİş* is indicating past tense/perfect aspect, or the evidential notions of inference or hearsay, and an epistemic statement with *-DİR* is indicating genericity or possibility depends on the linguistic context, that is, what other temporal-aspectual-modal meanings are expressed compositionally on the verb (Csató, 2000; Tura, 1986). The situational context is also of significance. For example, *-mİş* may serve a perceptive function expressing new information/surprise, as well as being the form for talking about the nonfactual realm as in narratives, play, and irony (Aksu-Koç et al., 2009; Johanson, 2000; Slobin & Aksu-Koç, 1982; Uzundağ et al., 2018). Similarly, the interpretation of *-DİR* changes depending on context; it may be used to convey an assumption in conversation or for an official announcement asserted with certainty (Aksu-Koç, 2016; Tura, 1986). In the present

and *-miş, miş, muş, müş*. Alternating vowels and consonants are represented by uppercase characters. We used the following abbreviations in glossing the examples: ACC: accusative, EVID: evidential, GEN: generic, PAST: past, PERF: perfect, POSS: possible.

study, we examine only the inferential and hearsay functions of *-mİş* and the certainty/generic function of *-Dİr*.

4. Reliability implications of Turkish evidentiality markers

Turkish-speaking adults associate direct evidentiality with high epistemic certainty and judge statements marked with *-Dİ* as more reliable compared to hearsay statements marked with *-mİş* (Arslan, 2020; Aydın, 2011; Aydın & Fitneva, 2019; Tosun & Vaid, 2018) and assumptions marked with *-mİş-Dİr* (Aydın, 2011; Aydın & Ceci, 2009; Tosun & Vaid, 2018). They also evaluate hearsay statements with *-mİş* and assumptions with *-mİş-Dİr*, both of which imply lack of perceptual evidence, as expressing lower certainty and lower reliability compared to inferential *-mİş* statements based on a resultant state that serves as perceptual evidence (Arslan, 2020; Tosun & Vaid, 2018). The privileged status of first-hand over second-hand sources also influences memory processes. Adults have stronger recall, recognition, and source memory for first-hand than for second-hand information which suggests that they tend to discard hearsay information expressed with *-mİş* as less trustworthy compared to first-hand information expressed with *-Dİ* (Aydın & Ceci, 2009; Aydın & Fitneva, 2019; Tosun et al., 2013).

Children's reliability attributions parallel those of adults. Experimental studies have revealed that four-year-olds judge information expressed from a direct experiencer perspective *-Dİ* to be more reliable than information based on hearsay *-mİş* and assumptions based on previous knowledge *-mİş-Dİr*; however, they do not differentiate the latter two sources in terms of reliability (Aydın, 2011). Four-to-five-year-olds judge direct evidential statements as being more certain than assumptions (Aksu-Koç & Alici, 2000), and five-to-seven-year-olds treat them as more reliable compared to both inferential and hearsay statements (Çelik et al., 2023; Öztürk & Papafragou, 2016). Suggestibility studies are also informative; four-to-six-year-olds resist misinformation expressed with the indirect evidential more than misinformation presented with the direct evidential and five-to-six-year-olds accept the suggestion of a direct witness regardless of their earlier perspectives (Aydın, 2011; Aydın & Ceci, 2009). In a similar vein, three-to-five-year-old children revise their beliefs more after directly observing an event than after indirectly witnessing or just hearing a report about it, again revealing higher reliability attribution to direct evidence than to the testimony of a witness or hearsay (Özkan et al., 2023).

Converging evidence, thus, shows that Turkish-speaking children understand the linguistic cues indicating reliability of information and judge statements of direct evidence to be more trustworthy and certain than statements expressing inference, hearsay, and assumptions around age four. However, we do not know whether children observe a reliability difference between inferential versus hearsay uses of the suffix *-mİş*. They acquire knowledge about the world from their own experience and from the linguistic reports, that is, the testimony of others. Inferential *-mİş* statements based on first-hand observation, even if partial, are likely to be regarded as more trustworthy than hearsay *-mİş* statements based on someone else's linguistic report. Understanding that different knowledge sources may differ in terms of reliability is critical because it affects whether the information is worth learning, keeping in memory, generalizing to other situations, sharing with others, or should be discarded as misinformation. We also have little information about their understanding of the uses of *-Dİr* for certain and generalizable knowledge, except for one study (Tamm et al., 2014), the findings of which have led to the questions investigated in the current study, as described below.

5. The current study

In Tamm et al.'s (2014) study, Turkish-learning four- and six-year-olds were found to generalize a property to other instances of the same kind when the property was communicated by use of the generic marker *-Dİr*, but not when expressed by use of a neutral predicate or the evidential marker *-mİş*. However, in this study, the evidential statement could be interpreted either as an inference or hearsay since the pragmatic contexts for differentiating the two functions of the indirect evidential were not provided. Tamm et al.'s findings, therefore, raise the question of why generalization is blocked when information is marked as acquired through indirect means. Do children interpret such knowledge as not generalizable because it is inferred from results of a non-witnessed process, or because it is based on hearsay, two different sources to which they may be attributing different degrees of reliability? In the current study, reasoning that generalization of knowledge rests on the ability to judge the reliability of its source, we modified Tamm et al.'s study of generalizability by differentiating the inference versus hearsay functions of the indirect evidential *-mİş*. In addition, we assessed children's reliability attributions to inferential versus hearsay uses of *-mİş* to see if a lower degree of trust is a factor that explains the lack of generalization of knowledge accessed indirectly. If, however, reliability does not make a difference to generalization of information presented in either the context of inference or that of hearsay, this would suggest that the use of the evidential *-mİş* form is restricted to particular situations, that is, to episodic information which is often about transient, specific, and independently verifiable events unlike semantic information that is generalizable and difficult to verify independently (Stephens & Koenig, 2015, pp. 182–183).

Thus, in the current study, we explore whether Turkish-speaking children's reliability attributions to evidential statements of inference versus hearsay are associated with their generalization of that information to instances of the same kind. Specifically, we investigate whether four- and six-year-olds (1) attribute different degrees of reliability to the uses of evidential *-mİş* as an inference versus a hearsay marker; (2) generalize information to other instances of a kind when expressed with (i) the evidential *-mİş* in contexts of inference from partial observable evidence, (ii) the evidential *-mİş* in contexts of hearsay, and (iii) the genericity marker *-Dİr*, as a baseline for generalization performance; and (3) whether their assessments of reliability of the two functions of the evidential marker are associated with their generalizing behaviour.

We expected children to attribute a higher degree of reliability to evidential information expressed in contexts of an inference from partial observable evidence, compared to evidential information expressed in hearsay contexts, and we expected six-year-olds to do so more than four-year-olds. This expectation was motivated by the different degrees of indirectivity associated with the two functions; observed results of a non-witnessed process constitute first-hand evidence even if partial, whereas hearsay lacks such direct evidential support. Second, we expected children to generalize the property of an object to other instances of the same kind when it is expressed with a statement marked with the generic *-Dİr* more than when it is expressed with a statement marked with *-mİş* indicating an inference or a statement marked with *-mİş* indicating hearsay, and we expected six-year-olds to show this tendency more strongly than 4-year-olds. This prediction was based on the fact that *-Dİr* is the generic marker in the language, and on Tamm et al.'s findings (2014) that both four- and six-year-olds have already grasped its generalizing function. Third, we expected children's reliability attributions to be related to their generalizations. Specifically, we expected children to generalize more in the case of the

inferential use of *-mİş* than the hearsay use of *-mİş* because they regard the former as more reliable.

6. Method

6.1. Participants

Forty-eight monolingual Turkish speaking four-year-olds ($M_{\text{months}} = 52.23$, $SD = 3.65$; 28 boys) and 48 six-year-olds ($M_{\text{months}} = 74.73$, $SD = 3.23$; 19 boys) participated in the study. This sample size is comparable to that of previous research on the acquisition of evidentiality in Turkish (Aksu-Koç et al., 2009; Ozturk & Papafragou, 2016; Ünal & Papafragou, 2016).

Participants were recruited from private kindergartens and private schools in upper-middle-class areas of two cities in Turkey and were tested individually in their schools. Ethical approval was obtained from the university review board for human subjects, and written consent forms were taken from the families. All families were native speakers of Turkish, and Turkish was the only language children were exposed to. Each participant’s parents completed a demographic information questionnaire that targeted information regarding the child’s date of birth, parents’ education, and income level. No significant difference was found between the age groups regarding parental education and household income (Table 1). Two participants were not included in the sample because they did not give codable data due to a lack of interest.

6.2. Measures

Evidentiary Reliability Task. This task was designed to assess children’s attributions of reliability to statements expressed with the evidential marker *-mİş* used in its inferential and hearsay functions. An introductory video plus seven animation videos were prepared in Adobe Flash Program and presented on a 13-inch MacBook Air laptop. The introduction video presented Ali and Ece, a brother and sister, living on a farm with their animals. Then a narrator’s voice said “There is a lot of mischief going on in this farm, and it is not well understood who creates it. There are three cute monsters who try to find out who is responsible for each mischievous event. Let’s see what is happening in this farm.” During this introduction, the experimenter pointed out some details such as Ali’s cap, Ece’s eyepatch, and some of the animals that figure as cues in the items (Figure 1).

There were two warm-up and five test trials. The scenarios were piloted for their difficulty levels and five scenarios that had similar levels of difficulty were chosen as test

Table 1. Demographic characteristics of participants

Variables	N	Four-year-olds M (SD)	N	Six-year-olds M (SD)	Test Statistics
Maternal education	46	15.65 (2.99)	45	16.13 (2.74)	$F(1,90) = .64, p = .43$
Paternal education	44	15.77 (3.12)	46	16.57 (2.72)	$F(1,89) = 1.66, p = .20$
Household income	43	3.53 (1.61)	35	3.46 (1.17)	$F(1,77) = .06, p = .81$

Note. Missing data are due to unanswered questions in the questionnaires. Maternal and paternal education represent the total years of education. Household income is categorized from 1 (low) to 6 (high)



Figure 1. The introductory scene for the video where the farm characters are presented.

trials. For each trial, the child saw a short scene representing an event described by one of the following change-of-state verbs: *ye* “eat,” *ısır* “bite,” *patlat* “pop,” *dök* “spill,” *kır* “break,” *iç* “drink,” and *yırt* “tear.” Each video-clip showed an initial state (e.g., a birthday cake on a table) and then a resultant state (e.g., one fourth of the cake is missing) indicating a change due to an unwitnessed past process (e.g., a slice of the cake had been eaten). Then two cute monsters who had not witnessed the past process appeared in two consecutive scenarios and made contrasting judgements about the agent of the event. These judgements were prerecorded as part of the video.

In the Hearsay scenario, a monster appears and says, *Aa, pasta yen-miş. Pastayı kim ye-miş?* “Oh, the cake has been eaten. Who ate the cake?” Another monster comes and whispers in the first monster’s ear who then reports what his friend told him, saying: *Pastayı Ali ye-miş* “Ali ate the cake, reportedly” (Figure 2). In the Inferential scenario, a monster appears and says, *Aa, pasta yen-miş. Pastayı kim ye-miş?* “Oh, the cake has been eaten. Who ate the cake?” and then refers to observable evidence (Ece’s pirate eye patch on the table) saying *Pastayı Ece ye-miş*. “Ece ate the cake, evidently/I infer” (Figure 3).

After watching the two scenarios, the child was asked by the experimenter to indicate which monster she thought was telling the truth: *Pastayı kim ye-miş?* “Who ate the cake?”. The appearance of the cute monsters across the inference and hearsay scenarios, and the

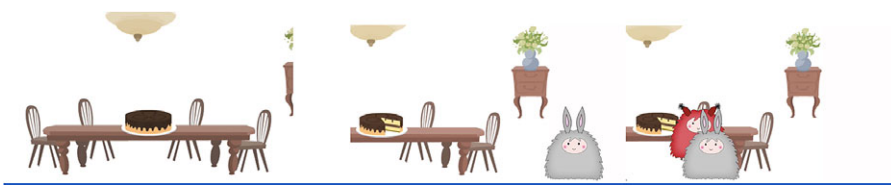


Figure 2. A Hearsay scenario. Grey monster: *Aa, pasta yen-miş. Pastayı kim ye-miş?* “Oh, the cake has been eaten. Who ate the cake?” Then red monster appears and whispers to the grey monster’s ear. Grey monster: *Pastayı Ali ye-miş* “Ali ate the cake, reportedly.”



Figure 3. An Inference scenario. Green monster: *Aa, pasta yen-miş. Pastayı kim ye-miş?* “Oh, the cake has been eaten. Who ate the cake?” Then the monster refers to observable evidence (Ece’s pirate eye patch on the table). Green monster: *Pastayı Ece ye-miş.* “Ece ate the cake, evidently/I infer.”

order of presentation of inference and hearsay scenarios across the task items were counterbalanced. All animations are provided in the [Supplementary Material](#).

Generalizability Task. This task, also used in the Tamm et al. (2014) study, is an adaptation of the “Blicket-Test” (Butler & Markman, 2012) and was designed to assess children’s generalizing behaviour. As testing materials, eleven 2.5 cm × 2.5 cm × 5 cm rectangular wooden blocks covered with green tape for 2/3 of their length and with black tape for 1/3 of their length were used as novel objects. All blocks were perceptually identical, but one of them had a magnet hidden in its black end (the “active block”), whereas the others did not (the “inert blocks”). The blocks were labelled “bilikit” in Turkish, after “blicket” in English.

Children were randomly assigned to three experimental conditions, generic, inferential, and hearsay, with 16 participants in each age group per condition. The generic condition which has been shown to lead to sustained exploratory behaviour in the Tamm et al. study (2014) was included to get a baseline measure of children’s generalizing behaviour. In all three conditions, the procedure was the same except for the stimulus sentence and the context of its presentation. In all three conditions, the experimenter told the children that she would teach them how to make a paper boat, but she first wanted to introduce a new toy, called “Bilikit,” showing and naming the wooden block with a sentence giving categorical information about the novel object. Children learned the label of the block and could successfully identify it among four distractor objects on two successive trials. Then, the experimenter put some paperclips on the table and demonstrated the property of “magnetism” using different everyday magnetic objects to pick up paperclips and encouraged the child to try the same. Thus, it was made sure that children knew what a “blicket” is and what being magnetic is. Then the magnetic objects were put away while the paper clips remained on the table. To compare the inferential and the hearsay uses of the evidential *-mİş* in terms of generalizability, we used slightly different scenarios without any ostensive or pedagogical demonstration. The contexts of presentation and the stimulus sentences in each condition were as follows:

1. Generic condition: The experimenter took the blicket in her hand, looked at it and said: *Bilikit mıknatıslı-dır*, “Blicket is magnetic [generic].”
2. Inferential condition: The experimenter took the blicket that already had a paperclip attached to its magnetic end in her hand, looked at it and said: *Bilikit mıknatıslı-ymış*, “Blicket is magnetic [evidently, as I infer].”
3. Hearsay condition: The experimenter’s cell phone rang, she answered it and looked at the blicket that she was holding while talking, and after hanging up the phone, she said: *Bilikit mıknatıslı-ymış*, “Blicket is magnetic [reportedly].”

In each condition, after the presentation of the stimulus sentence, the child was presented with 10 inert blickets and the paperclips on the table to play with while the experimenter would be looking for some paper to make the paper boat. The child who did not know that the blickets were not magnetic was free to play with the blickets for 60 seconds during which they were videotaped. The task has the advantage of eliciting an implicit behavioural response (children's exploratory behaviour) instead of a verbal response that requires explication of knowledge.

6.3. Procedure

Children were tested individually in a quiet room in their schools by the first author. The tasks were administered in the same order to all participants. For each child, the experiment started with one of the three conditions of the Generalizability task and continued with the Evidentiary Reliability Task. 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.

Data Coding and Inter-Rater Reliability (IRR). The data were coded by the first author. IRR was established by an independent trained coder who coded a randomly selected 25% of the videos for each task. The IRR scores were analysed on SPSS version 28 (IBM Corp., Armonk, NY, USA) and are reported below for each task separately.

Evidentiary reliability task scoring

- (i) *Inference reliability percentage.* Children's choices of inference from observable evidence (rather than hearsay) were coded as "1," and these scores were added up across trials, leading to scores between 0 and 5. As one child left one question unanswered, the inference reliability percentage was calculated, and the total number of children's inference reliability choices (0–5) was divided by the total trials that they answered (4–5). For example, if a child chose the inferential source as more reliable than the hearsay source 3 times over 5 trials, the inference reliability percentage was calculated as .60 for this participant. Cohen's kappa = 1.00.

Generalizability task scoring. Three scores were calculated and used as dependent variables in the analyses (Tamm et al., 2014):

- (i) *Duration (in seconds).* This score refers to the time spent on trying the inert blickets for magnetism during the first 60 seconds (i.e., how many seconds the child spent on trying the blickets for magnetism, ranging between 0–60).
- (ii) *Number of blickets.* This score refers to the total number of inert blickets the child tried for magnetism during the first 60 seconds (i.e., how many different blickets the child tried for magnetism, ranging between 0 and 11).
- (iii) *Total number of trials.* This score refers to the total number of trials that the child tested the inert blickets for magnetism during the first 60 seconds (i.e., how many times the child tried the blickets for magnetism). For example, one participant may try 1 blicket for magnetism 10 times, and another participant may try 10 different blickets once. Total number of trials will be 10 for both participants, yet the number of blickets will be different (1 for the first participant and 10 for the second participant).

Cohen's kappa for the three variables (duration = .99, number of blickets = 1.00, and total number of trials = .90) indicates almost perfect agreement between the coders.

7. Results

All variables were checked for the assumptions of the parametric analyses used. The normal distribution assumption was violated for the generalizability task variables (i.e., duration, number of blickets, total trials). Also, skewness and kurtosis values were not within acceptable limits ($+ - 1.96$ SD), and the residuals of the generalizability task variables were not normally distributed. Therefore, logarithmically transformed scores were used for these variables (See Table 2 for descriptive statistics of untransformed scores).

7.1. Children's assessment of the reliability of inferential and hearsay statements

To test whether both age groups reliably chose inference over hearsay, we compared children's scores to chance with one sample t-tests. Results showed that both four- and six-year-olds chose inference over hearsay above chance level (Chance = 0.5, $t(47) = 4.85$, $p < .001$, $d = .70$, and $t(47) = 9.46$, $p < .001$, $d = 1.37$), respectively.

Children's inference reliability percentage scores were also submitted to a one-way ANOVA with age (4 years, 6 years) as the between-subjects variable. The results showed that six-year-olds endorsed statements based on inference ($M = .81$, $SD = .23$) significantly more than four-year-olds ($M = .64$, $SD = .20$), $F(1,94) = 15.64$, $p < .001$, $\eta^2 = .14$ (Figure 4).

Table 2. Descriptive statistics for reliability and generalizability scores by age

Variables		Age Group			
		Four-year-olds ($N = 48$)		Six-year-olds ($N = 48$)	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Reliability (%)	Inferential - <i>mlş</i>	0.64	0.20	0.81	0.23
Generalizability	<i>Duration in seconds</i>				
	Generic - <i>Dlr</i>	32.06	24.77	32.50	21.31
	Inferential - <i>mlş</i>	6.19	15.02	23.00	18.27
	Hearsay - <i>mlş</i>	5.56	9.14	21.25	19.64
	<i>Number of Blickets</i>				
	Generic - <i>Dlr</i>	2.44	2.45	2.88	2.34
	Inferential - <i>mlş</i>	0.69	1.25	2.37	2.45
	Hearsay - <i>mlş</i>	0.69	1.01	1.63	0.96
	<i>Total Trials</i>				
	Generic - <i>Dlr</i>	10.63	10.88	11.87	10.33
	Inferential - <i>mlş</i>	2.00	5.95	7.63	6.49
	Hearsay - <i>mlş</i>	1.31	1.96	7.31	9.03

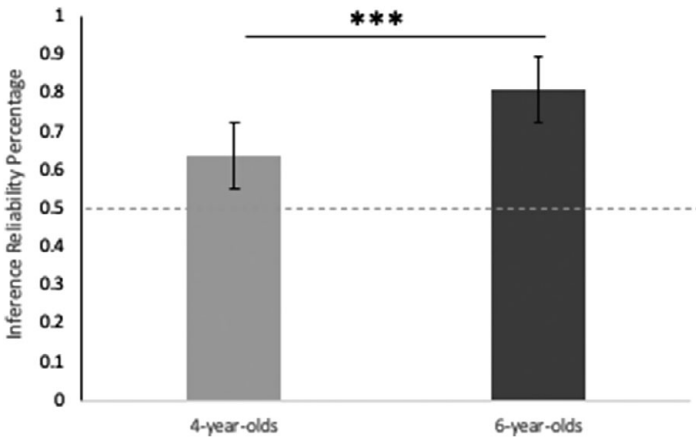


Figure 4. Children's mean reliability attributions (%) to inferential sources by age. Note. Error bars represent standard error, and the dashed line represents the chance level (0.5). *** $p < .001$.

7.2. Children's generalization performance in response to generic, inferential, and hearsay statements

A two-way MANOVA was conducted on children's generalization scores (duration of trying, number of blickets tried, total number of trials) as dependent variables with linguistic condition (generic *-Dlr*, inferential *-mIs*, hearsay *-mIs*) and age (4 years, 6 years) as between-subject variables. The multivariate test revealed a significant effect of linguistic condition, Wilks' Lambda (Λ) = .82, $F = 3.04$, $p = .007$, $\eta_p^2 = .09$, and a significant effect of age, Wilks' Lambda (Λ) = .78, $F = 8.09$, $p < .001$, $\eta_p^2 = .22$ on generalization scores. No significant interaction was found between age and linguistic condition on generalization scores, Wilks' Lambda (Λ) = .92, $F = 1.31$, $p = .25$, $\eta_p^2 = .04$.

Test of between-subjects showed significant effects of linguistic condition on the duration of time children tried the blickets (Figure 5), $F(2, 90) = 9.09$, $p < .001$, $\eta_p^2 = .17$, on the total number of trials they carried out (Figure 6), $F(2, 90) = 9.16$, $p < .001$, $\eta_p^2 = .17$, and on the number of blickets they tried (Figure 7), $F(2, 90) = 7.14$, $p = .001$, $\eta_p^2 = .14$, to test the magnetic property of the blickets. Bonferroni comparisons demonstrated that children in the baseline generic *-Dlr* condition tried blickets for longer duration of time ($M = 1.30$, $SD = .59$) than children in the inferential *-mIs* ($M = .79$, $SD = .68$), $p = .001$, and the hearsay *-mIs* ($M = .78$, $SD = .65$), $p = .001$, conditions. Children in the generic *-Dlr* condition also tried a higher number of blickets ($M = .48$, $SD = .28$) than children in the inferential *-mIs* ($M = .30$, $SD = .29$), $p < .01$, and the hearsay *-mIs* ($M = .28$, $SD = .23$), $p < .01$, conditions. The total number of trials was also higher in the generic *-Dlr* condition ($M = .88$, $SD = .49$) than the inferential *-mIs* ($M = .50$, $SD = .48$), $p < .01$, and the hearsay *-mIs* ($M = .48$, $SD = .45$), $p = .001$, conditions. However, differences between inferential and hearsay *-mIs* conditions on duration, number of blickets, and total number of trials were not significant ($ps > .99$).

The effect of age on children's generalization behaviour was significant. Six-year-olds tried blickets for longer duration of time $F(1, 90) = 24.69$, $p < .001$, $\eta_p^2 = .22$, tried a higher number of blickets $F(1, 90) = 17.27$, $p < .001$, $\eta_p^2 = .16$, and made more trials $F(1, 90) = 20.74$, $p < .001$, $\eta_p^2 = .19$ than four-year-olds. In addition, there was a significant interaction between age and linguistic condition on duration of trying,

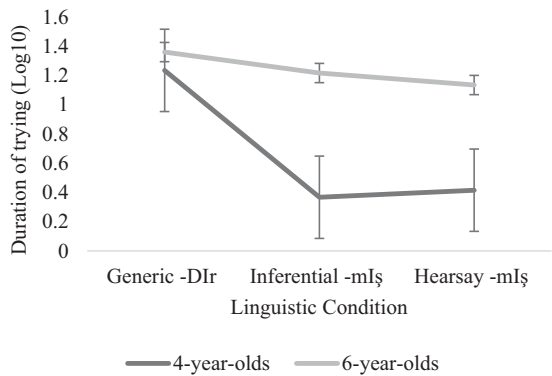


Figure 5. Mean duration of trying by age and linguistic condition. *Note.* Error bars represent standard error.

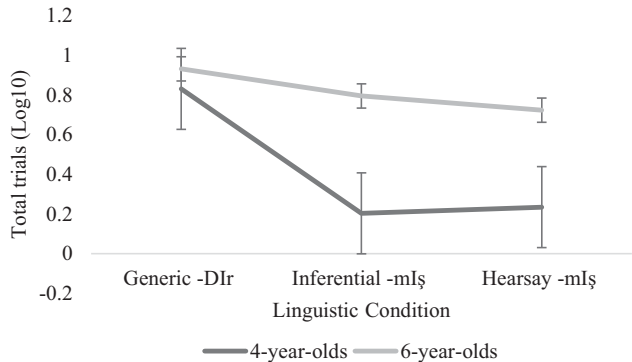


Figure 6. Mean total number of trials by age and linguistic condition. *Note.* Error bars represent standard error.

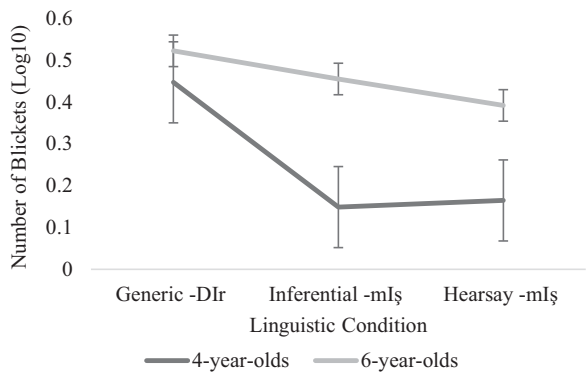


Figure 7. Mean number of blickets tried by age and linguistic condition. *Note.* Error bars represent standard error.

$F(2, 90) = 3.85, p = .025, \eta_p^2 = .08$ (Figure 5), and a marginally significant interaction on the number of total trials, $F(2, 90) = 2.99, p = .055, \eta_p^2 = .06$ (Figure 6). However, there was no significant interaction between age and linguistic condition on the number of blickets children tried, $F(2, 90) = 1.93, p = .15, \eta_p^2 = .04$ (Figure 7).

Bonferroni-adjusted simple main effect tests were computed for the significant interactions (see Table 2 for means and standard deviations). Results showed that four- and six-year-olds' duration of trying did not significantly differ from each other in the generic *-DIr* condition ($p = .53$). However, six-year-olds' duration of trying was higher than four-year-olds' in the inferential *-mIş* condition ($p < .001$) and in the hearsay *-mIş* condition ($p < .001$). No difference was found in six-year-olds' duration of trying between the three linguistic conditions ($ps > .77$). However, four-year-olds tried the blickets for magnetism for a longer time in the generic *-DIr* condition than the inferential *-mIş* condition ($p < .001$) and hearsay *-mIş* condition ($p < .001$). No difference was found in four-year-olds' duration of trying between the inferential and the hearsay *-mIş* conditions ($p = 1.00$) (Figure 5).

Bonferroni-adjusted simple main effect tests demonstrated a similar pattern with the total number of trials. Four- and six-year-olds' total number of trials was not different in the generic *-DIr* condition ($p = .50$). Yet, six-year-olds tried the blickets more often than four-year-olds in the inferential *-mIş* condition ($p < .001$) and in the hearsay *-mIş* condition ($p = .002$). No difference was found in 6-year-olds' total number of trials between the three linguistic conditions ($ps > .50$). However, four-year-olds tried blickets more often in the generic *-DIr* condition than in the inferential *-mIş* condition ($p < .001$) and in the hearsay *-mIş* condition ($p < .001$). There was no difference in four-year-olds' total number of trials between the inferential and the hearsay *-mIş* conditions ($p = 1.00$) (Figure 6).

7.3. Relationship between reliability and generalizability

To examine whether the reliability of information source predicts children's generalizing performance, linear regression analyses were computed with inference reliability percentage scores as the predictor (independent variable) and the logarithmically transformed scores of duration, number of blickets, and total number of trials as outcomes (dependent variables) for inference and hearsay linguistic conditions. Since age has a significant effect on generalizability performance, linear regression analyses were computed for the two age groups separately (see Table 3).

Results did not reveal a significant relationship between children's reliability attributions to inferential source (over hearsay) and their generalization performance in either the inferential *-mIş* or hearsay *-mIş* conditions (Table 3).

8. Discussion

The current study investigated Turkish-speaking children's reliability attributions to grammaticalized markers of evidential source and whether reliability associated with different degrees of indirectivity influences the generalization of that knowledge. Specifically, we examined whether four-to-six-year-olds regard evidential statements with *-mIş* used in inferential contexts as more reliable than statements with *-mIş* used in hearsay contexts. Then we asked whether they generalize information expressed with a *-mIş* statement used in an inferential context more than information expressed with a *-mIş*

Table 3. Linear regression results for predicting generalization (duration, number of blickets, total trials) by inference reliability attributions

	Four-year-olds						Six-year-olds					
	<i>N</i>	<i>R</i> ²	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>N</i>	<i>R</i> ²	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>
<i>Inferential -mİş</i>	16						16					
Duration		.12	.95	.68	1.39	.19		.03	.40	.58	.69	.50
Number of Blickets		.13	.41	.28	1.46	.17		.06	-.28	.30	-.93	.37
Total trials		.12	.61	.44	1.39	.19		.02	.27	.49	.55	.59
<i>Hearsay -mİş</i>	16						16					
Duration		.06	.81	.89	.91	.38		.00	.03	.55	.05	.96
Number of Blickets		.04	.27	.35	.77	.45		.05	-.14	.17	-.84	.42
Total trials		.04	.38	.50	.76	.46		.01	.13	.46	.30	.77

Note. Log¹⁰ transformations were used for the duration, number of blickets, and total trials variables.

statement in a hearsay context. We also assessed children's responses to generic *-Dir* statements for their generalizing behaviour. Finally, we asked whether children's assessments of the reliability of the two functions (inference vs hearsay) of the evidential marker *-mİş* predict their generalizing behaviour.

8.1. Reliability of inferential and hearsay statements

Our first expectation stated that children would attribute a higher degree of reliability to evidential information expressed in contexts of an inference from partial observable evidence compared to evidential information expressed in hearsay contexts that lack any observable evidence. The results confirmed our prediction; both four- and six-year-olds chose inferential sources as more reliable than hearsay sources. Six-year-olds chose inference as more reliable compared to hearsay more than four-year-olds, as predicted. The difference between the two age groups suggests that understanding the reliability of information inferred from partial observable evidence compared to second-hand information advances between four to six years and accords with Butler et al.'s (2020) suggestion that children's understanding of source reliability continues to develop up to six years of age.

Previous research has demonstrated that children are sensitive to abstract linguistic cues indicating the reliability of types of information sources and degrees of epistemicity. English-speaking three-to-four-year-olds understand the implied reliability of modal and mental verbs that express epistemic gradations of certainty expressed lexically (Diessel & Tomasello, 2001; Hirst & Weil, 1982; Leahy & Zalnieriunas, 2021; Moore et al., 1989; Shatz et al., 1983) around the same age as children speaking languages that mark these notions morphologically. Japanese, Bulgarian, Korean, and Turkish four-to-six-year-olds find first-hand information expressed with direct evidential markers more reliable than information expressed with inference and hearsay markers (Aydin, 2011; Çelik et al., 2023; Fitneva, 2008; Matsui et al., 2006; Matsui & Miura, 2009; Özkan et al., 2023; Ozturk & Papafragou, 2016; Papafragou et al., 2007). The current results additionally

demonstrate that young children can use grammatical cues and pragmatic context to make more nuanced evaluations of source reliability, differentiating between inference and hearsay. This order, whereby full or partial first-hand access to information ranks above hearsay, lends support to the idea that ownership of information impacts the certainty and reliability of information conveyed (Arslan, 2020, p.12; Tosun & Vaid, 2018, p.153).

8.2. Generalizability of generic, inferential, and hearsay statements

Our second expectation concerned children's generalizing behaviour based on linguistic information in the absence of an ostensive demonstration. As predicted, children of both ages attributed generalizability to statements with *-DIr*. Our results confirmed Tamm et al.'s (2014) findings that the suffix *-DIr* has already been evaluated as a generic marker by age four. This is in line with previous findings in English indicating four-year-old children's tendency to generalize information more when generic than non-generic sentences are used in the absence of any ostensive or pedagogical demonstration (Chambers et al., 2008; Cimpian & Scott, 2012).

However, contrary to our expectation, children did not generalize differentially in response to inferential versus hearsay statements with *-mİş*, but different patterns were observed for the two age groups. Four-year-olds engaged in equally low generalizing behaviour in response to both the inferential and hearsay uses of *-mİş*, whereas six-year-olds sustained a significantly high level of exploratory behaviour, trying the blickets for their magnetic property almost equally for either function.

When children's generalization patterns to inferential and hearsay uses of *-mİş* were compared to their generalizations in response to a generic statement with *-DIr*, again different patterns were observed for four- and six-year-olds. As predicted, four-year-olds generalized more when information was expressed with the generic marker *-DIr* than when expressed with either the inferential or the hearsay uses of the evidential *-mİş* in response to which they generalized minimally. However, six-year-olds differed, generalizing almost equally in response to statements marked with *-DIr* for genericity, to statements with *-mİş* in inferential contexts and *-mİş* in hearsay contexts.

The performance of our four-year-olds shows that they have already differentiated the functions of the evidential *-mİş* and generic *-DIr* markers in terms of generalizability. Their behaviour is similar to that of children who, in suggestibility studies, tend to discard the information expressed with *-mİş* as less trustworthy compared to the direct evidential form (Aydin, 2011; Aydin & Ceci, 2009). These results are in accord with Tamm et al.'s findings for four-year-olds but not for six-year-olds. In contrast to Tamm et al. (2014), our six-year-olds' generalization behaviour did not significantly differ in response to *-DIr*, inferential *-mİş*, and hearsay *-mİş* statements. Several explanations come to mind for their unexpected generalizing behaviour. First, children's representation of the functions of the evidential *-mİş* and generic *-DIr* may be changing with age. Previous experimental studies have yielded ages three to four years for differentiated production of evidential markers, and five (Aksu-Koç, 1988; Aksu-Koç et al., 2009) or even six years of age (Ozturk & Papafragou, 2016; Ünal & Papafragou, 2016) for comprehension. Six-year-olds may be testing and reorganizing the boundaries of the semantics and pragmatics of these multifunctional suffixes at the interface of the evidential and epistemic domains in terms of the types of evidence and speaker stance they encode in different contexts (Aksu-Koç, 2016). Second, their behaviour brings to mind an early acquired function of *-mİş*

expressing the speaker's cognitive realization that the knowledge just received is "new" information (Slobin & Aksu-Koç, 1982), and therefore worth exploring further. Third, six-year-olds may be generalizing more because they want to see and verify for themselves that the blickets, a novel object category, are indeed magnetic, similar to Butler et al.'s (2020) six-year-olds who attributed more reliability to a source who verified their claims regardless of their previous reliability. Fourth, six-year-olds, already attending school, may have been hypothesizing about the expectations in the testing situation and trying to meet the demands made by the experimenter possibly regarded as a trustworthy teacher.

8.3. Reliability and generalizability

Our third expectation concerned whether information from the evidential source judged as more reliable would be generalized more than information obtained from the source evaluated as less reliable. Although both our four- and six-year-olds attributed higher reliability to inference over hearsay as evidential source, they did not generalize information more in response to inferential *-mİş* statements compared to hearsay *-mİş* statements. Four-year-olds made equally low generalizations in the case of inferential and hearsay statements, whereas six-year-olds made equally high generalizations in case of both, regardless of their reliability judgements. Thus, the findings run counter to our hypothesis predicting higher generalization in the case of inferential than hearsay *-mİş* statements because children would regard the former use more trustworthy than the latter. That is, the lack of differential generalization to evidential statements cannot be explained by the different degrees of reliability associated with inference versus hearsay interpretations. Instead, they indicate that reliability is not the (only) factor blocking the generalization of information conveyed by indirect evidential statements.

An alternative explanation is that the use of the indirect evidential *-mİş* is restricted to communicating specific, episodic events, either inferred from a particular piece of evidence or acquired as second-hand information about a particular situation, while *-Dİr* is the form in the language to use for statements of generic force. The performance of our four-year-olds and that of the four- and six-year-olds of Tamm et al.'s (2014) study suggests that children have differentiated these functions of the evidential *-mİş* and the generic *-Dİr* suffixes already at age four. Sensitivity to differences between specific versus generic knowledge has also been evidenced by English-speaking three- and four-year-olds who preferred an informant with a previous history of reliability about semantic information to an informant with a previous history of reliability about episodic information (Stephens & Koenig, 2015). This interpretation does not hold for our six-year-olds, possibly because their understanding of the functions of evidential *-mİş* as appropriate to specific, episodic information was undermined by their motivation to verify the blicket's magnetic property, as was discussed above. Another reason might be that our limited sample size restricted the power of the statistical analyses. Future studies are needed to explore this relationship further.

8.4. Contributions and limitations

The current study is the first to integrate the question of source reliability and generic knowledge acquisition by exploring it via the use of the grammaticalized markers of these notions in Turkish in the same research. To the best of our knowledge, this is also the first

study that differentiates the inferential and hearsay functions of the evidential marker *-mİş*, investigating their implications for reliability and generalizability.

Our methodological contributions are (1) modifying the generalization task to differentiate the inferential and hearsay functions of the evidential marker *-mİş*, and (2) developing a new task to assess the source reliability of the evidential *-mİş* based on contextual differences implying its two different functions, and (3) using tasks that do not require explicit verbalization. In the generalization task, children gave actional, that is, procedural responses, and in the reliability task, they only had to make a choice but were not required to produce an explanation. Our theoretical contribution is to show that the Turkish indirect evidential marker *-mİş* indeed has strong epistemic implications (Aksu-Koç, 2016; Johanson, 2000; Terziyan & Aksu-Koç, 2021), as evidenced by the fact that even children as young as four years of age attribute different degrees of reliability to its inferential and hearsay uses.

As for limitations, it could be argued that in the inference scenarios of the reliability task children have perceptual access to the evidence themselves, leading to a comparison of the hearsay statement to their own inference which could have resulted in higher reliability attributions than to hearsay. While this remains a possibility, the two conditions were closely matched with the cute monsters appearing in both types of scenarios to ensure that children took both into account in making their judgements. Furthermore, the inference scenario videos did not offer immediately perceivable evidence for the children to pick up (except in the case of paw prints signalling a cat as the agent). Children, therefore, had to rely on the linguistic message of the monsters to make their judgements.

The present findings suggest that language structure does not make a difference for understanding the reliability implications of epistemic and evidential constructions, whether grammaticized or lexical, as research with English-speaking children also indicates age four for the corresponding epistemic distinctions based on reliability (Diessel & Tomasello, 2001; Hirst & Weil, 1982; Leahy & Zalnieriunas, 2021; Moore et al., 1989; Shatz et al., 1983). However, in order to come to a decisive conclusion on this issue and to further expand the present findings, it is necessary to carry out controlled crosslinguistic research with languages that have obligatory evidential markers, as well as those that mark the same notions lexically, and also to include three-year-olds to the comparison.

9. Conclusion

Several studies have investigated the influence of behavioural and linguistic cues on children's understanding of generic knowledge (Csibra & Gergely, 2009; Prasada, 2000) and learning based on the reliable testimony of others (Harris et al., 2018; Koenig & Sabbagh, 2013). Here, we provide evidence from Turkish, where a distinction between direct and indirect sources of information as well as knowledge genericity is grammaticalized in morphology. Children acquire knowledge from adults through language, and our results show that by 4 years of age, Turkish-speaking children understand the linguistic cues that signal types of information sources, degrees of reliability, and generalizable versus non-generalizable knowledge. The findings also suggest that these abilities show further development between 4 and 6 years of age that involve some reorganization of the boundaries of evidential and epistemic indicators and their relations. The exact nature of these reorganizations, however, needs to be explored in future studies.

Acknowledgments. We thank Çağla Aydın, Aylin Küntay, and Ayşecan Boduroğlu for their feedback during the conception of this study and Treysi Terziyan and Stefani Terziyan for their help with preparing animations.

Author contribution. This article was produced from Merve Ataman-Devrim's master's thesis, which was completed under the supervision of Ayhan Aksu-Koç and Gaye Soley at Boğaziçi University. Ethics approval was obtained from the institutional review board at Boğaziçi University (SBB-EAK 2017/8). Anonymized data and the analysis outputs are shared on Open Science Framework (https://osf.io/r6dxw/?view_only=18272c7e24274711b668f82f326d250f). This study was not preregistered. Merve Ataman-Devrim served as lead for data collection, formal analysis, visualization, and writing-original draft. Ayhan Aksu-Koç and Gaye Soley served as lead for supervision, resources, writing-review, and editing and served in a supporting role for formal analysis. The authors equally contributed to conceptualization and methodology of the study.

Funding statement. This research did not receive any funding. Gaye Soley is funded by the Serra Hunter Programme, Generalitat de Catalunya.

Competing interests. The authors declare none.

Disclosure of use of AI tools. The authors declare none.

References

- Aboudy, R., Yousif, S. R., Sheskin, M., & Keil, F. C. (2022). Says who? Children consider informants' sources when deciding whom to believe. *Journal of Experimental Psychology: General*, **151**(10), 2481–2493. <https://doi.org/10.1037/xge0001198>.
- Aksu-Koç, A. (1988). *The acquisition of aspect and modality. The case of past reference in Turkish*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511554353>.
- 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 honor of Eser Erguvanlı-Taylan, studies in language companion series* (Vol. 175, pp. 143–156). John Benjamins B.V.
- Aksu-Koç, A., & Alica, D. M. (2000). Understanding sources of beliefs and marking of uncertainty: The child's theory of evidentiality. In *Proceedings of the 30th annual child language conference* (pp. 123–130). Stanford University.
- Aksu-Koç, A., Ögel-Balaban, H., & Alp, İ. E. (2009). Evidentials and source knowledge in Turkish. In S. A. Fitneva & T. Matsui (Eds.), *Evidentiality: A window into language and cognitive development, new directions for child and adolescent development* (Vol. 125, pp. 13–28). Jossey-Bass. <https://doi.org/10.1002/cd.247>.
- Arslan, S. (2020). When the owner of information is unsure: Epistemic uncertainty influences evidentiality processing in Turkish. *Lingua*, **247**, 1–16.
- Aydın, C. (2011). *Language and suggestibility: Cross-linguistic evidence on children's assessment of source knowledge*. [Doctoral thesis, Cornell University]. <https://hdl.handle.net/1813/30620>
- Aydın, Ç., & Ceci, S. J. (2009). Evidentiality and suggestibility: A new research venue. In S. A. Fitneva & T. Matsui (Eds.), *Evidentiality: A window into language and cognitive development, new directions for child and adolescent development* (Vol. 125, pp. 79–93). Jossey-Bass. <https://doi.org/10.1002/cd.251>.
- Aydın, Ç., & Fitneva, S. (2019). Türkçe'deki geçmiş zaman bildirme kiplerinin tümce hatırlamaya etkisi. *Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi*, **59**(1), 353–369. <https://doi.org/10.33171/dtcjournal.2019.59.1.19>.
- Butler, L. P., Gibbs, H. M., & Tavassolie, N. S. (2020). Children's developing understanding that even reliable sources need to verify their claims. *Cognitive Development*, **54**, 100871. <https://doi.org/10.1016/j.cogdev.2020.100871>.
- Butler, L. P., & Markman, E. M. (2012). Preschoolers use intentional and pedagogical cues to guide inductive inferences and exploration: Pedagogical cues and exploration. *Child Development*, **83**(4), 1416–1428. <https://doi.org/10.1111/j.1467-8624.2012.01775.x>.
- Butler, L. P., Schmidt, M. F. H., Bürgel, J., & Tomasello, M. (2015). Young children use pedagogical cues to modulate the strength of normative inferences. *British Journal of Developmental Psychology*, **33**(4), 476–488. <https://doi.org/10.1111/bjdp.12108>.

- Butler, L. P., & Tomasello, M. (2016). Two- and 3-year-olds integrate linguistic and pedagogical cues in guiding inductive generalization and exploration. *Journal of Experimental Child Psychology*, **145**, 64–78. <https://doi.org/10.1016/j.jecp.2015.12.001>.
- Buttelmann, D., & Zmyj, N. (2020). Fourteen-month-olds' imitation is influenced more strongly by a model's competence than by a model's certainty. *Infant Behavior and Development*, **60**, 101458. <https://doi.org/10.1016/j.infbeh.2020.101458>.
- 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. <https://doi.org/10.1111/j.1467-8624.2012.01860.x>.
- Çelik, B., Ergut, N., & Allen, J. W. (2023). What's the evidence say? The relation between evidential-trust and theory of mind. *Journal of Cognition and Development*, 1–20. <https://doi.org/10.1080/15248372.2023.2260874>.
- Chambers, C. G., Graham, S. A., & Turner, J. N. (2008). When hearsay trumps evidence: How generic language guides preschoolers' inferences about unfamiliar things. *Language and Cognitive Processes*, **23**(5), 749–766. <https://doi.org/10.1080/01690960701786111>.
- Choi, S. (2021). The development of sentence-ending epistemic/evidential markers in young Korean children. In U. Stephany & A. Aksu-Koç (Eds.), *Development of modality in first language acquisition: A cross-linguistic perspective* (pp. 491–524). Walter de Gruyter GmbH & Co KG.
- Cimpian, A., & Scott, R. M. (2012). Children expect generic knowledge to be widely shared. *Cognition*, **123**(3), 419–433.
- Clement, F., Koenig, M., & Harris, P. (2004). The ontogenesis of trust. *Mind and Language*, **19**(4), 360–379. <https://doi.org/10.1111/j.0268-1064.2004.00263.x>.
- Corriveau, K.H., Pickard, K. & Haris, P.L. 2010. Preschoolers trust particular informants when learning new names and new morphological forms. *British Journal of Developmental Psychology*, **29**(1), 46–63. <https://doi.org/10.1348/2044-835X.002009>
- Csató, É. Á. (2000). Turkish MiŞ- and iMiŞ-items. Dimensions of a functional analysis. In L. Johanson & B. Utas (Eds.), *Evidentials: Turkic, Iranian and neighbouring languages* (pp. 29–43). Mouton de Gruyter.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, **13**(4), 148–153. <https://doi.org/10.1016/j.tics.2009.01.005>.
- Diessel, H., & Tomasello, M. (2001). The acquisition of finite complement clauses in English: A corpus-based analysis. *Cognitive Linguistics*, **12**(2), 97–142. <https://doi.org/10.1515/cogl.12.2.97>.
- Fitneva, S. A. (2008). The role of evidentiality in Bulgarian children's reliability judgments. *Journal of Child Language*, **35**(4), 845–868. <https://doi.org/10.1017/S0305000908008799>.
- Gelman, S. A., & Raman, L. (2003). Preschool children use linguistic form class and pragmatic cues to interpret generics. *Child Development*, **74**(1), 308–325. <https://doi.org/10.1111/1467-8624.00537>.
- Harris, P. L., Koenig, M. A., Corriveau, K. H., & Jaswal, V. K. (2018). Cognitive foundations of learning from testimony. *Annual Review of Psychology*, **69**(1), 251–273. <https://doi.org/10.1146/annurev-psych-122216-011710>.
- Hirst, W., & Weil, J. (1982). Acquisition of epistemic and deontic meaning of modals. *Journal of Child Language*, **9**(3), 659–666. <https://doi.org/10.1017/S0305000900004967>.
- Johanson, L. (2000). Turkic indirectives. In L. Johanson & B. Utas (Eds.), *Evidentials: Turkic, Iranian and neighbouring languages* (pp. 61–87). Mouton de Gruyter.
- 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. <https://doi.org/10.1111/j.0956-7976.2004.00742.x>.
- Koenig, M. A., Cole, C. A., Meyer, M., Ridge, K. E., Kushnir, T., & Gelman, S. A. (2015). Reasoning about knowledge: Children's evaluations of generality and verifiability. *Cognitive Psychology*, **83**, 22–39. <https://doi.org/10.1016/j.cogpsych.2015.08.007>.
- Koenig, M. A., & Harris, P. L. (2005a). Preschoolers mistrust ignorant and inaccurate speakers. *Child Development*, **76**(6), 1261–1277. <https://doi.org/10.1111/j.1467-8624.2005.00849.x>.
- Koenig, M. A., & Harris, P. L. (2005b). The role of social cognition in early trust. *Trends in Cognitive Sciences*, **9**(10), 457–459. <https://doi.org/10.1016/j.tics.2005.08.006>.
- Koenig, M. A., & Sabbagh, M. A. (2013). Selective social learning: New perspectives on learning from others. *Developmental Psychology*, **49**(3), 399–403. <https://doi.org/10.1037/a0031619>.
- Leahy, B., & Zalnieriunas, E. (2021). Might and might not: Children's conceptual development and the acquisition of modal verbs. *Semantics and Linguistic Theory*, 426–445. <https://doi.org/10.3765/salt.v31i0.5082>.

- Matsui, T., & Fitneva, S. A. (2009). Knowing how we know: Evidentiality and cognitive development. In S. A. Fitneva & T. Matsui (Eds.), *Evidentiality: A window into language and cognitive development, new directions for child and adolescent development* (Vol. 125, pp. 1–11). Jossey-Bass. <https://doi.org/10.1002/cd.246>.
- Matsui, T., & Miura, Y. (2009). Children's understanding of certainty and evidentiality: Advantage of grammaticalized forms over lexical alternatives. In S. A. Fitneva & T. Matsui (Eds.), *Evidentiality: A window into language and cognitive development, new directions for child and adolescent development* (Vol. 125, pp. 63–77). Jossey-Bass. <https://doi.org/10.1002/cd.250>.
- 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. <https://doi.org/10.1016/j.cogdev.2005.10.001>.
- Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development*, 60(1), 167–171. <https://doi.org/10.2307/1131082>.
- Novack, I. A., Ho, S., & Sera, M. (1996). Children's understanding of epistemic modals. *Journal of Child Language*, 23(3), 621–643. <https://doi.org/10.1017/S0305000900008977>.
- Özkan, F. E., Hartwell, K., & Köymen, B. (2023). A cross-linguistic approach to children's reasoning: Turkish- and English-speaking children's use of metatalk. *Developmental Science*, e13374. <https://doi.org/10.1111/desc.13374>.
- Ozturk, O., & Papafragou, A. (2016). The Acquisition of Evidentiality and Source Monitoring. *Language Learning and Development*, 12(2), 199–230. <https://doi.org/10.1080/15475441.2015.1024834>.
- Palmer, F. R. (2001). *Mood and modality*. Cambridge University Press.
- Papafragou, A., Li, P., Choi, Y., & Han, C. (2007). Evidentiality in language and cognition. *Cognition*, 103(2), 253–299.
- Poulin-Dubois, D., Brooker, I., & Polonia, A. (2011). Infants prefer to imitate a reliable person. *Infant Behavior and Development*, 34(2), 303–309. <https://doi.org/10.1016/j.infbeh.2011.01.006>.
- Prasada, S. (2000). Acquiring generic knowledge. *Trends in Cognitive Sciences*, 4(2), 66–72. [https://doi.org/10.1016/S1364-6613\(99\)01429-1](https://doi.org/10.1016/S1364-6613(99)01429-1).
- Robinson, E. J., Champion, H., & Mitchell, P. (1999). Children's ability to infer utterance veracity from speaker informedness. *Developmental Psychology*, 35(2), 535.
- Scofield, J., & Behrend, D. A. (2008). Learning words from reliable and unreliable speakers. *Cognitive Development*, 23(2), 278–290. <https://doi.org/10.1016/j.cogdev.2008.01.003>.
- Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition*, 14(3), 301–321.
- Slobin, D. I., & Aksu-Koç, A. (1982). Tense, aspect and modality in the use of the Turkish evidential. In P. J. Hopper (Ed.), *Tense-aspect between semantics and pragmatics* (pp. 185–200). John Benjamins B.V.
- Stephens, E. C., & Koenig, M. A. (2015). Varieties of testimony: Children's selective learning in semantic versus episodic domains. *Cognition*, 137, 182–188.
- Tamm, A., Çağlar, L. R., Aksu-Koç, A., & Csibra, G. (2014). *Generic grammar makes Turkish preschoolers generalize statements and tolerate exceptions* [Poster presentation]. *The Budapest CEU conference on cognitive development (BCCCD 2014)*, Budapest, Hungary.
- Taylan, E. E. (2001). On the relation between temporal/aspectual adverbs and the verb form in Turkish. In E. E. Taylan (Ed.), *The verb in Turkish* (pp. 97–128). John Benjamins.
- Terziyan, T., & Aksu-Koç, A. (2021). Epistemic and evidential modality in early Turkish child speech. In U. Stephany & A. Aksu-Koç (Eds.), *Development of modality in first language acquisition: A cross-linguistic perspective* (pp. 453–490). Walter de Gruyter GmbH & Co KG.
- Tosun, S., & Vaid, J. (2018). Source vs. stance: On the relationship between evidential and modal expressions. *Dialogue & Discourse*, 9(1), 128–162.
- Tosun, S., Vaid, J., & Geraci, L. (2013). Does obligatory linguistic marking of source of evidence affect source memory? A Turkish/English investigation. *Journal of Memory and Language*, 69(2), 121–134.
- Tura, S. S. (1986). DİR in modern Turkish. In A. Aksu-Koç & E. Erguvanlı-Taylan (Eds.), *The proceedings of the Turkish linguistic conference* (pp. 145–158). Boğaziçi University Publications.
- Ünal, E., & Papafragou, A. (2016). Production–comprehension asymmetries and the acquisition of evidential morphology. *Journal of Memory and Language*, 89, 179–199. <https://doi.org/10.1016/j.jml.2015.12.001>.

- Uzundağ, B. A., Taşçı, S. S., Küntay, A. C., & Aksu-Koç, A.** (2018). Functions of Turkish evidentials in early child–caregiver interactions: A growth curve analysis of longitudinal data. *Journal of Child Language*, 45(4), 878–899. <https://doi.org/10.1017/S0305000917000526>.
- 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. <https://doi.org/10.1016/j.jecp.2010.03.002>.

Cite this article: Ataman-Devrim, M., Soley, G., & Aksu-Koç, A. (2025). Children's Understanding of Source Reliability and Knowledge Generalizability from Grammatical Cues: Evidence from Turkish. *Journal of Child Language* 1–23, <https://doi.org/10.1017/S0305000925100421>