Deafness & Education International

ISSN: 1464-3154 (Print) 1557-069X (Online) Journal homepage: http://www.tandfonline.com/loi/ydei20

Using the theory of mind inventory to detect a broad range of theory of mind challenges in children with hearing loss: a pilot study

Tiffany L. Hutchins, Lyndsey Allen & Maggie Schefer

To cite this article: Tiffany L. Hutchins, Lyndsey Allen & Maggie Schefer (2017): Using the theory of mind inventory to detect a broad range of theory of mind challenges in children with hearing loss: a pilot study, Deafness & Education International, DOI: 10.1080/14643154.2016.1274089

To link to this article: http://dx.doi.org/10.1080/14643154.2016.1274089

Published online: 23 Jan 2017.

DEAFNESS & EDUCATION INTERNATIONAL, 2017
http://dx.doi.org/10.1080/14643154.2016.1274089

Using the theory of mind inventory to detect a broad range of theory of mind challenges in children with hearing loss: a pilot study

Tiffany L. Hutchins[a], Lyndsey Allen[b] and Maggie Schefer[a]

[a]University of Vermont, Burlington, VT, USA; [b]The Ear Foundation, Nottingham, UK

ABSTRACT

Traditional child-performance measures of theory of mind (ToM) are associated with several limitations. The Theory of Mind Inventory-2 (ToMI-2) is a new broadband caregiver-informant measure designed to tap children's ToM competence. The purposes of this pilot study were to (1) gather preliminary data to explore the scope of the ToM challenges experienced by oral and late-signing children with hearing loss (CHL) and, (2) gather pilot data to explore the criterion-related validity of the ToMI-2 for use with this population. ToMI-2 results revealed wide variation in ToM strengths and challenges both within and between individuals. ToMI-2 scores also positively correlated with scores for hearing vocabulary and a test of pragmatic language development and negatively correlated with the age of cochlear implantation. The present results are encouraging for the use of the ToMI-2 to detect ToM challenges in CHL. Clinical implications and directions for future research are discussed.

ARTICLE HISTORY
Received 3 November 2016
Accepted 14 December 2016

KEYWORDS
Social cognition; theory of mind; assessment; validity

Once thought to be uniquely disrupted in autism, much research now demonstrates that Theory of Mind (ToM) challenges are evidenced in a wide range of clinical populations including oral and late-signing children and adolescents with hearing loss (CHL). In the first (self-described) study on this topic, Peterson and Siegal (1995) tested a group of prelingually deaf children (ages 8–13 years) of hearing parents on a series of standard false belief tasks (described more fully below). They found that 65% of children failed the tasks which are routinely passed by hearing children between 4- and 5-years of age. Peterson and Siegal (1995) concluded that the ToM deficits of CHL rivalled those of autism although the processes that mediate ToM challenges in autism (i.e. a core conceptual deficit) are understood to be different than those involved in hearing loss (i.e. a conversational deficit).

In the two decades since Peterson and Siegal's (1995) groundbreaking study, much research has affirmed their findings and conclusions while fleshing out a more comprehensive portrait of the nature and correlates of ToM challenges in CHL. Although not universally observed, one frequently reported correlate of ToM abilities is language ability (e.g. Macaulay and Ford 2006, Schick *et al.* 2007, Pyers and Senghas 2009, de Villiers and de Villiers 2014) which, in turn, is entangled awith socialization. Indeed, there is wide variability in the age at which CHL succeed on ToM tasks with much of this variation attributed to the diversity of the samples in terms of hearing history, social experience, and the amount and quality of opportunities to observe and participate in conversation (e.g. Woolfe *et al.* 2002, Gonzalez *et al.* 2007, Pyers and de Villiers 2013). Of course, deafness and hearing loss ultimately represent a problem of access to language and social information and so it is not surprising that age of cochlear implantation is also negatively correlated with scores on measures of ToM. In a recent study, Sundqvist *et al.* (2014) compared an early CI group and a late CI group and found that the early group performed better on ToM tasks despite the fact that the groups were matched on chronological age and receptive vocabulary.

Variability in ToM in CHL has also been attributed to different methods of assessment (Pyers and de Villiers 2013) which is a primary focus of this paper. In the following section, we discuss (1) some traditional methods of assessing ToM in CHL (2) the limitations associated with these methods, (3) parents as experts of children's knowledge and abilities and, (4) our data analytic plan for a pilot study exploring the criterion-related validity of a new caregiver-informant measure of ToM (The Theory of Mind Inventory). It will ultimately be argued that a sound informant-measure of ToM for detecting a broader range of ToM challenges in CHL is long overdue.

Traditional measures of assessing ToM in CHL

Questions about the nature, timing, and manner in which ToM knowledge emerge have generated a variety of ToM assessment methods (with most being adapted from the autism literature). Assessment

CONTACT Tiffany L. Hutchins Tiffany.hutchins@med.uvm.edu University of Vermont, Burlington, VT, USA

measures range from tasks designed to tap CHL's understanding of desires, emotions, empathy, beliefs, belief-desire reasoning, or psychological explanation to assessments to the production and comprehension of mental state terms (e.g. 'want', 'think', 'know') (e.g. Gray *et al.* 2007, Peters *et al.* 2009, Hao and Chan 2010, Netten *et al.* 2015). Among the ToM assessment procedures that have been devised, the classic false belief task (Wimmer and Perner 1983) warrants special attention because of its prominence in research, practice, and theory for assessing typically developing children and children with a variety of clinical conditions including hearing loss. In this task, children are told a story in which an object is moved from an old location to a new location without the knowledge of the main protagonist.[1] For example, Sally puts a marble in a basket and leaves the room. In her absence, Anne enters and moves the marble from the basket to a box and then she leaves. Children are asked, 'When Sally returns, where she will look for the marble?' Children who answer with the new (incorrect) location fail the question whereas children who answer with the old (correct) location pass the question by presumably demonstrating their knowledge that behaviours are guided by inner mental states, in this case a false belief.

One major limitation of the use of the false belief task in isolation is that it taps only one aspect of a much more complex and multifaceted ToM. For this reason, some researchers have opted for the use of aggregate measures in the form of ToM scales to assess different components of ToM across levels of complexity. The rationale for this is that a broader range of tasks allows for the examination of the developmental sequence of discrete ToM abilities. By far, the measure that has had the most significant impact on research on the ToM abilities of CHL is the ToM Scale (Wellman and Lui 2004). The ToM Scale was developed on the results of meta-analytic contrasts that revealed a sequence in the development of five ToM abilities in typically developing preschoolers as follows:[2] Diverse Desires (understanding that two persons can have different desires about the same objects), Diverse Beliefs (understanding that two people can have different beliefs about the same object), Knowledge Access (what we call 'seeing-leads-to-knowing'; the understanding that seeing-leads-to knowing and not seeing leads to ignorance), False Belief (understanding that someone can hold a belief that contradicts reality), and Hidden Emotion (what we call 'display rules'; the understanding that a person can feel one emotion but display a different emotion). More recently, a sixth component of ToM (i.e. Sarcasm) was added and found to be the most advanced task on the scale (Peterson *et al.* 2012). The 5- and 6-step versions of the ToM Scale have been used by Peterson and colleagues (e.g. Peterson *et al.* 2005, Peterson and Wellman 2009, Peterson *et al.* 2012) in a series of studies over the last decade. Its use for evaluating the ToM in CHL is second only to the use of the classic false belief task used in isolation (but which is also represented on the ToM Scale). In short, the ToM Scale has immeasurably advanced our understanding of ToM in CHL and has played a pivotal role in research. Nonetheless, limitations of direct measures of ToM performance, like the standard false belief task and ToM Scale, should be considered in light of how these limitations might be overcome by the use of a complimentary or alternative ToM assessment methodology.

Limitations of traditional ToM assessment methods

Language and cognitive performance factors

An often cited potential limitation of traditional ToM tasks is that performance is complicated by cognitive and linguistic factors. Several researchers have examined the influence of language factors through the use of low- and nonverbal ToM tasks and have concluded that, although lessening the verbal demands of the tasks can improve the performance of all children (Figueras-Costa and Harris 2001), the gap between hearing and CHL in ToM performance remain. Thus, it appears that the comparatively poor performance of CHL on verbally administered tasks reflects a conceptual deficit as opposed to a receptive or expressive language deficit (de Villiers and de Villiers 2000, Figueras-Costa and Harris 2001, Woolfe *et al.* 2002). Whether performance on these tests reliably improves when linguistic demands are attenuated ignores three serious remaining problems. First, the influence of performance factors cannot be entirely eliminated in tasks that inherently require attention, memory, and understanding of the language involved. Second, very young children cannot even be tested with most traditional ToM tasks, since they lack the cognitive and verbal skills necessary to answer the control questions, success on which is usually an inclusion criterion. In fact, many researchers of ToM in CHL employ false belief tasks only when the individual can demonstrate a minimum verbal mental age of three to four years (e.g. Macauley and Ford 2006); thereby excluding very young children from participation (and the possibility of early risk detection). Third, although the group studies reviewed here vary in ToM task administration procedures, considerations for eliciting children's best and most reliable performance is typically formalized through the use of standard experimental protocols (and these are typically articulated in the method section of empirical works). In the applied world, more variability in elicitation procedures will occur because the children tested will represent a more heterogeneous group for which different testing accommodations might be appropriate and because of variability in resources (e.g. availability of fluent-signing interpreters; Pyers and de Villiers 2013). These and other factors will impact performance (increase error) and create the potential for seriously flawed interpretation. In light of all of these challenges, a reliable, valid, and standardized informant-measure of ToM for very young children as well as older individuals with limited verbal capacities would be valuable in research and clinical work.

Content validity

Peterson and colleagues use of the ToM Scale represents an important advancement in ToM research and it is well-suited for describing a developmental progression of a limited number of specific ToM competencies. Nonetheless, what is presently missing is broadband assessment of ToM. Even task batteries (which are designed to tap ToM domains beyond false beliefs) are narrow in scope considering the tremendous breadth of the construct. For example, additional areas of ToM include (but are by no means limited to): the understanding of (or ability to engage in) humour, pretence, deception, empathy, certainty, emotion recognition, joint attention and intentionality, the mental–physical distinction, the appearance-reality distinction, the causes and consequences of thoughts and emotions, first- and second-order thinking, visual perspective-taking, and complex social judgment (Hutchins *et al.* 2008, Hutchins *et al.* 2012, Hutchins and Prelock 2016). Moreover, ToM is tightly knotted up with (and perhaps at the heart of) the understanding of speech acts, pragmatic discourse, narrative construction, and metalinguistic aspects of language which can, thusly, be included in our conceptualization of ToM (Most *et al.* 2010, Peterson and Slaughter 2006). Given that ToM is a term that 'refuses to be corralled' (Astington and Baird 2005, p. 4) and that it is now routinely used, more or less interchangeably, with the terms 'social cognition', 'perspective-taking', and 'metacognition', it is clear that even existing ToM task batteries are extremely limited in their ability to capture the breadth of the content domain relevant to construct of ToM (Hutchins *et al.* 2008).

Ecological and social validity

Although performance on traditional ToM tasks have been shown to correlate with socially competent real-world behaviour in samples of children with CHL (de Rosnay *et al.* 2014), it is important to note that the explicit nature of traditional ToM tasks (e.g. 'Sally did not see that the ball was moved. Where will Sally look for her ball first?') does not resemble the ways that real life social dilemmas are presented. 'Not only are social demands in naturalistic settings not explicitly formulated as a problem-solving situation, they need to be created and defined as a "social demand" by the person' (Klin 2000, p. 832). Thus, individuals who pass a dichotomous false belief task or who perform well on the ToM Scale may be artificially credited with ToM competence, when in fact there is a continuum of competence that is revealed in daily social dysfunction (Klin 2000). A measure that is sensitive to fine variations in ToM and that relies on information accrued over time during real-world social interaction helps to move us toward assessment of ToM that is socially valid, that adopts a dimensional approach of social cognitive abilities (Klin 2000), and that may reveal ToM challenges (and strengths) that may be masked by traditional ToM tasks.

Parents as experts of their children's abilities

Involving caregivers as informants and interpreters of their children's behaviours is important because it reflects the growing recognition that caregivers possess expert knowledge regarding their children's abilities, strengths, and weaknesses and, as such, are reliable and invaluable sources of information (e.g. Crais 1993). Indeed, moderate to high correlations have been observed between parents' and professionals' judgments of a CHL's mental health strengths and challenges (Cornes and Brown 2012). This suggests that caregiver reports can yield valid indices of child functioning. In addition, many informant measures (which might be criticized on the basis of their potential for subjectivity and bias) have endured the scrutiny of rigorous psychometric evaluation (e.g. the MacArthur Communication Development Inventory; Fenson *et al.* 2007) thereby demonstrating construct validity and the accuracy of the information source.

It makes sense that parents (or other primary caregivers) would be experts about their children's abilities including their ToM competence. During social interaction, parents can accumulate rich information about the child's mind and develop accurate insights into the child's perspectives. Parents also have numerous opportunities to observe child ToM knowledge as it is applied (or not applied) in a range of real-world contexts. No current measures of ToM functioning take advantage of the knowledge of those who are closest to child. By contrast, our proposed measure is family centered and relies on the familiarity of adults who know the child best (McCauley 2001).

Data analytic plan

The purpose of this study was twofold. First, we wanted to explore the scope of ToM challenges in CHL using a broadband measure of ToM (i.e. the ToMI-2; described below). To address this goal, we used simple descriptive analyses where we quantify the percent of children for whom caregivers indicated item scores in the clinical range (defined as -1SD). At times, the constructs tapped by the ToMI-2 are the same as those tapped by the ToM Scale or other child-performance tasks. When ToMI-2 items corresponded to ToM domains for which researchers had previously reported fail rates for CHL (or in cases where these data could be calculated from reported data), these data were compared.

Because this pilot study is the first to evaluate the utility of the ToMI-2 for a sample of CHL, we sought preliminary data to inform the tool's potential criterion-related validity for this population. Given the links from ToM to language and pragmatic functioning, we first correlated composite ToMI-2 scores with results from a hearing vocabulary test and a caregiver-informant measure of pragmatics. If the ToMI-2 is a valid indicator of ToM in CHL, we would expect to find significant correlations. In light of the conversation hypothesis of ToM challenges in CHL, we also explored the relationship between scores on the ToMI-2 and age at first amplification and age at first implantation with the expectation that ToMI scores would be negatively correlated with these variables. Given our small sample size and in an effort to avoid type II errors, we chose an alpha of .10 for all analyses.

Method

Participants

Participants were 12 primary caregivers and their children (eight males, four females; ages 5 years 2 mo — 11 years 1 mo, $M = 8.5$ years). Children represented a range vis-à-vis hearing loss aetiology, age when hearing loss was discovered, age at first hearing aid fitting, and age at first cochlear implant. With regard to concomitant difficulties, children presented with a range of conditions including medical problems, vestibular difficulties, and learning disability. No children were included in this sample who had been diagnosed with a concomitant developmental disability (e.g. autism spectrum disorder, attention disorders) as these are known to be associated with language and ToM challenges in their own right.

On the basis of caregiver report, all children had sensorineural hearing loss. For all but one child, the degree of uncorrected hearing loss was characterized as profound (91 dB+) and the degree of corrected hearing loss was characterized as mild (26–40 dB). For the aforementioned child, hearing loss was characterized as moderate to severe (56–70) when uncorrected and as moderate (41–55 dB) when corrected. For the vast majority of children, the modal communication strategy at home and school was spoken language. More specific data for each of the 12 participants are presented in Table 1.

Measures

The British picture vocabulary scale — 3rd Edition (BPVS-3)

The BPVS-3 (Dunn *et al.* 2009) is a normed referenced wide-range test of hearing vocabulary. The third version of the BPVS is directly linked to the Peabody Picture Vocabulary Test (Dunn and Dunn 2007) and has been extensively normed and validated using a large stratified sample.

The pragmatic language skills inventory (PLSI)

The PLSI (Gilliam and Miller 2006) is a caregiver-informant measure designed to identify children (ages 5–12) who have pragmatic language difficulties. The PLSI consists of 45 items that comprise three subscales (each with 15 items). These are: (1) the Personal Interaction Skills subscale (e.g. initiating conversation, asking for help, participating in verbal games, and using appropriate nonverbal gestures), (2) the Social Interaction Skills (e.g. when to talk and when to listen, understanding classroom rules, taking turns in conversations, and predicting consequences), and (3) the Classroom Interaction Skills (e.g. using figurative language, maintaining topic, explaining how things work, writing stories, and using slang). The PLSI also yields on overall composite index.

Theory of mind inventory-2 (ToMI-2)

The ToMI-2 (Hutchins and Prelock 2016) is a revision of two previous iterations (i.e. the ToMI, Hutchins *et al.* 2012; the Perceptions of Children's ToM Measure — Experimental Version, Hutchins *et al.* 2008). All versions of the ToMI have been submitted to rigorous psychometric scrutiny by the developers as well as independent researchers. The measure performs extremely well under all tests of reliability and validity including: test-retest reliability for typically developing samples and an autism spectrum disorder sample for both short and long lags, internal consistency (i.e. Cronbach's alpha), multiple tests of criterion-related validity, contrasting-groups validity, and factor analytic tests of construct validity (Hutchins *et al.* 2008, 2012, Lerner *et al.* 2011, Greenslade and Coggins 2016).

The ToMI-2 is a caregiver-informant broadband measure designed to tap a wide range of social cognitive understandings. Each item takes the form of a statement (e.g. 'My child understands whether someone hurts another on purpose or by accident') and was developed to serve as a face valid indicator of a particular dimension of ToM. Each of the 60 items comprising the ToMI-2 belong to one of three empirically derived subscales (i.e. Early, Basic, and Advanced) that reflect a developmental progression in ToM. The Early items represent ToM domains that are mastered in typical development in late infancy and toddlerhood, the Basic items reflect ToM

Table 1. Participant characteristics.

Child age (years)	Gender	Age (weeks) at first HA	Age (weeks) at first CI	HL aetiology	Concomitant difficulties	Modal comm. Strategy: home	Modal comm. Strategy: school	BPVS-3 SS (%)	PLSISS (%)	ToMI-2: Composite SS (%)	ToMI-2: Early T-score (%)	ToMI-2: Basic T-score (%)	ToMI-2: Advanced T-score (%)
5.17	M	10	67	Cytomegalovirus	None	Spoken language	Spoken language	114 (82%)	96 (39%)	90 (30%)	36 (15%)	54 (59%)	40 (19%)
9.50	M	8	95	Unknown	None	Spoken language	Spoken language supported by sign	< 70 (< 1st)	92 (30%)	75 (1%)	48 (47%)	28 (1%)	26 (1%)
9.34	M	65	82	Meningitis	Vestibular difficulties	Spoken language	Spoken language	106 (66%)	106 (65%)	89 (33%)	47 (41%)	43 (25%)	42 (23%)
8.67	M	13	n/a	Connexin 26	Learning disability	Spoken language	Spoken language	71 (<2%)	90 (25th)	65 (1%)	23 (1%)	26 (1%)	21 (1%)
10.08	F	n/a	130	Unknown	None	Spoken language	Spoken language	82 (12%)	116 (86%)	75 (8%)	50 (44%)	34 (11%)	35 (6%)
8.25	M	26	78	Unknown	None	Spoken language	Spoken language supported by sign	74 (4%)	94 (35%)	68 (3%)	32 (7%)	27 (1%)	31 (5%)
10.08	M	95	147	Unknown	None	Spoken language	Spoken language	112 (78%)	106 (65%)	59 (5%)	35 (11%)	29 (1%)	28 (6%)
11.08	F	13	61	Connexin 26	None	Spoken language	Spoken language	106 (66%)	98 (45%)	70 (1%)	50 (39%)	26 (1%)	29 (1%)
7.50	M	9	43	Connexin 26	Vestibular difficulties	Spoken language	Spoken language	93 (32%)	94 (35%)	88 (30%)	30 (1%)	50 (47%)	43 (18%)
9.17	M	17	65	Unknown	Hirschprung's disease	Spoken language	Spoken language	115 (84%)	120 (91%)	97 (39%)	55 (60%)	50 (41%)	42 (21%)
5.84	F	9	100	Usher syndrome type 1B	Vestibular difficulties; prog. vision loss	Spoken language	Spoken language	110 (74%)	92 (30%)	101 (47%)	53 (55%)	51 (46%)	50 (46%)
10.50	F	78	290	Auditory neuropathy spectrum disorder	Learning disability, dyslexia, dyscalculia	Spoken language supported by sign	Spoken language supported by sign	< 70 (< 1$^{s\%}$)	59 (< 1$^{s\%}$)	49 (< 1$^{s\%}$)	29 (< 1$^{s\%}$)	29 (< 1$^{s\%}$)	18 (< 1$^{s\%}$)

Notes: BPVS-3 = British Picture Vocabulary Scale-3; PLSI = Pragmatic Language Skills Test; ToMI-2 = Theory of Mind Inventory-2: standard scores on the composite score: M = 100, SD 15; standard scores on the Early, Basic, and Advanced subscales are T-scores: M = 50, SD = 10.

domains mastered in the preschool years, and Advanced items reflect ToM domains mastered in late childhood. Three additional rationally derived subscales (i.e. Emotion Recognition, Mental State Term Comprehension, Pragmatics) are also available. The ToMI-2 was normed on a large and ethnically diverse U.S. sample (Hutchins and Prelock 2016).

Procedure

The parent and child participants in this study were recruited following parent's registration with The Ear Foundation research forum where they expressed a willingness to take part in research. All data were collected by a RCSLT registered Speech and Language Therapist who administered the test of receptive vocabulary and distributed questionnaires to the primary caregivers. All data were collected in the families' homes during a single visit. Research conducted by The Ear Foundation follows guidelines laid out by the British Education Research Association (2011) and procedures developed with the University of Nottingham. All research are subject to internal and external ethics approval and adheres to Data Protection regulation

Results

Descriptive statistics

Table 1 presents the composite standard scores for the BPVS-3, PLSI, and ToMI-2. Inspection of this table reveals that 7 of the 12 children (i.e. 58.3%) in this pilot study evidenced receptive vocabulary in the normal range (i.e. standard score > 85), 11/12 (91.6%) children evidenced scores on the PLSI that were in the normal range, and 5/12 (41.6%) of children evidenced scores on the ToMI-2 in the normal range. With regard to ToMI-2 subscale scores (which are T-scores where $M = 50$; $SD = 10$), an almost universal trend was noted where the Early subscale items showed the greatest strength, followed by the Basic subscale items, and the lowest scores were evidenced for the Advanced items.

As composite scores can obscure deficits at the item-level (and because this is a particular concern for those CHL who may be generally high functioning but for whom subtle deficits may be detectable within a specific ToM domain), we also explored the scope of ToM challenges in CHL for all children across all ToMI-2 items. Using simple descriptive analyses, we quantified the percent of children for whom caregivers indicated item scores in the clinical range (a.k.a. 'fail rates' as defined as the percent of children for whom an item score was -1SD). Where corresponding ToM domains had been assessed in previous studies using measures of direct child performance, those data are also reported (see Table 2).

Inferential statistics

Our preliminary investigation explored correlations between composite ToMI-2 scores with BPVS-3 and the PLSI with the expectation that a valid indicator of ToM for CHL would show significant correlations. Pearson's product moment correlations revealed a significant correlation between the ToMI-2 composite and BPVT-3 standard scores ($r = .58$, $p = .081$) with hearing vocabulary explaining approximately 34% of the

Table 2. ToMI-2 item domains by subscale and the number (and %) of individuals with that domain disrupted (i.e. scores – 1 SD).

Early subscale: dimension intended to be tapped	ToMI-2 fail rates: % cases where domain disrupted	Fail rates (% of cases) in which domain was disrupted as reported in previous studies using direct tests of child performance
Early empathy	33.0	
Discrimination of basic emotions	16.6	
Intentionality	33.0	
Basic positive emotion recognition (happy)	25.0	
Social referencing: reading fear	41.6	
Sharing attention: initiating	0.0	
Sharing attention: responding	8.3	
Gaze following	41.6	
Social referencing: ambiguous situation	8.3	
Basic negative emotion recognition (sad)	25.0	
Basic negative emotion recognition (mad)	16.6	
Basic negative emotion recognition (scared)	16.6	
Mental state term comprehension: early desire (want)	8.3	
Desire-based emotion	8.3	6% (Peterson *et al.* 2012) 9% (Peterson and Wellman 2009) 8% (Peterson *et al.* 2005) 7% for young children with CI; 0% for older children with CI (Remmel and Peters 2008)
Basic subscale: dimension intended to be tapped		
Physiologically based behaviour	33.3	
Emotion-based behaviour	16.6	
Mental state term comprehension: cognitive terms (think)	25.0	

(*Continued*)

Table 2. Continued.

Early subscale: dimension intended to be tapped	ToMI-2 fail rates: % cases where domain disrupted	Fail rates (% of cases) in which domain was disrupted as reported in previous studies using direct tests of child performance
False beliefs: unexpected location	50.0	48% (Peterson *et al.* 2012) 71% (Peterson and Wellman 2009) 20% (Macauley and Ford 2006) 67% (Peterson *et al.* 2005) 79% for young children with CI; 13% for older children with CI (Remmel and Peters 2008) 65% (Peterson and Seigal 1995)
Seeing-leads-to-knowing	58.3	36% Peterson *et al.* 2012) 58% (Peterson and Wellman 2009) 47% (Peterson *et al.* 2005) 60% for young children with CI; 0% for older children with CI (Remmel and Peters 2008)
Mental state term comprehension: cognitive terms (know)	25.0	
Appearance-reality distinction	41.6	50% (Macauley and Ford 2006)
False beliefs: unexpected contents	66.6	48% (Peterson *et al.* 2012) 71% (Peterson and Wellman 2009) 70% (Macauley and Ford 2006) 67% (Peterson *et al.* 2005) 79% for young children with CI; 13% for older children with CI (Remmel and Peters 2008) 65% (Peterson and Seigal 1995)
Certainty	33.0	
Mental–physical distinction	41.6	
Pretense: engaging in pretense	16.6	
Counterfactual reasoning	33.0	
Pretense: understanding pretense in others	16.6	
Child's ability to deceive	41.6	
Cognitive emotion recognition (disgust)	33.3	
Speech acts: performatives (promises)	33.3	
Pragmatics: secrets	16.6	
Mental state term comprehension: cognitive terms (belief)	8.3	
Attribute-based behaviour	41.6	
Cognitive emotion recognition (surprise)	25.0	
Mental state term comprehension: desire (need)	25.0	
Duture thinking (self)	8.3	
Belief-based emotion	16.6	
Advanced subscale: dimension intended to be tapped		
Pragmatics: sarcasm	66.6	97% (Peterson *et al.* 2012)
Second-order false desire attribution	50.0	
Pragmatics: idiomatic language	75.0	
Pragmatics: deception by others	66.6	
Emotion recognition: display rules	75.0	81% (Peterson *et al.* 2012) 80% (Peterson and Wellman 2009) 72% (Peterson *et al.* 2005) 57% for young children with CI; 7% for older children with CI (Remmel and Peters 2008)
Pragmatics: complex social judgment (understanding bullying vs. teasing)	58.3	
Pragmatics: white lies	66.6	
Pragmatics: lies versus jokes	58.3	
Visual perspective-taking (level 2)	33.3	50% (Falkman *et al.* 2007)
Second-order understanding of belief	58.3	
Second-order understanding of emotion	41.6	
Complex social judgment (understanding when speaking partner not interested)	41.6	
True empathy	33.3	
Pragmatics: humour (play on words)	66.6	
Biased cognition	50.0	
Ambiguous figure perception	66.6	
Pragmatics: audience adaptation	25.0	
Mixed emotions	75.0	
Common sense: social knowledge	41.6	
Complex emotion recognition (embarrassed)	41.6	
Complex emotion recognition (guilt)	25.0	
Autonoetic (self-knowing) awareness	8.3	
Mental state inference in context	16.6	

variance in ToMI-2 scores. We also found a significant correlation between the PLSI and ToMI-2 composite standard scores ($r = .51$, $p = .093$) with each measure sharing approximately 26% variance with the other.

In light of the conversation hypothesis of ToM challenges in CHL, we also explored the relationship between scores on the ToMI-2 and age at first amplification and age at first implantation with the expectation that ToMI scores would be negatively correlated with these variables. With the exception of the Emotion Recognition subscale (which is comprised more heavily of Early subscale items), all ToMI-2 subscales correlated with hearing vocabulary. The ToMI-2 composite score also strongly negatively correlated with age of amplification ($r = -.63$, $p = .03$) and age of implantation ($r = -.75$, $p = .008$) as did a number of other ToMI-2 subscales. With regard to age of amplification versus age of implantation, the directions of the correlations with ToMI-2 scores (composite and subscales) where the same although the correlations for implantation were uniformly stronger (and this was true for the PLSI as well). The full correlation matrix for subscale and composite scores for the ToMI-2 and PLSI are presented in Table 3.

Discussion

The first purpose of this study was to explore the scope of ToM challenges in CHL using a broadband measure of ToM. As inspection of Table 1 indicates, the children in this study were diverse in terms of their HL aetiology, hearing vocabulary, pragmatic development, and age at amplification and implantation. Not surprisingly, the children were also diverse in ToM abilities as measured by the ToMI-2 with scores ranging from the clinical to age-typical range and a trend was also noted such that the greatest ToM strengths tended to be more numerous among Early developing ToM domains, less frequent for Basic domains, and least apparent for Advanced domains.

When ToM deficits were reported by caregivers (operationalized in this study as scores falling at least 1 SD below the mean), the diversity of deficits across ToM skill areas was particularly striking and no area other than 'initiating joint attention' appeared universally spared in our small sample. Still, some areas appeared more frequently affected: as ToMI-2 items increased from Early, to Basic, to Advanced, data for the CHL indicated more limited development. Finally, when ToMI-2 data for 'fail rates' were compared to data for studies that examined these same ToM competencies using the method of direct child testing, ToMI-2 data were very much in line with previous reports, (often around the median percent reported in previous studies), thus offering more evidence in support of the ToMI-2's construct validity.

The results from this pilot study are important from both research and clinical perspectives. The ToMI-2 can be used to identify specific challenge areas that can be targeted for intervention but identifying challenge areas is only the first step in treatment target selection. A critical second step is to determine whether and which treatment targets are also developmentally appropriate. It is surprising how often this is overlooked in interventions to support ToM. One significant

Table 3. Correlation matrix (α = .10) for receptive vocabulary, PLSI, and ToMI-2 subscale and composite scores with age at amplification/implantation.

Variable	BPVS-3	PLSI: PIS	PLSI: SIS	PLSI: CI	PLSI: Composite	ToMI-2: Early	ToMI-2: Basic	ToMI-2: Advanced	ToMI-2:MS Terms	ToMI-2:Emo. Recog.	ToMI-2:Pragmatics	ToMI-2:Composite
BPVS-3	–	$r = -.16$	$r = .32$	$r = .51$	$r = .30$	$r = .56$	$r = .56$	$r = .57$	$r = .59$	$r = .44$	$r = .58$	$r = .58$
	–	$p = .62$	$p = .37$	$p = .13$	$p = .40$	$p = .09$[a]	$p = .09$[a]	$p = .08$[a]	$p = .07$[a]	$p = .19$	$p = .08$[a]	$p = .08$[a]
Age first HA	$r = .23$	$r = -.25$	$r = -.19$	$r = -.11$	$r = -.16$	$r = -.23$	$r = -.32$	$r = -.41$	$r = -.45$	$r = -.17$	$r = -.63$	$r = -.55$
	$p = .55$	$p = .45$	$p = .59$	$p = .75$	$p = .64$	$p = .49$	$p = .34$	$p = .21$	$p = .15$	$p = .61$	$p = .037$*	$p = .07$[a]
Age first CI	$r = -.04$	$r = -.67$	$r = -.68$	$r = -.48$	$r = -.63$	$r = -.36$	$r = -.43$	$r = -.77$	$r = -.71$	$r = -.44$	$r = -.75$	$r = -.70$
	$p = .92$	$p = .022$*	$p = .021$*	$p = .13$	$p = .035$*	$p = .28$	$p = .19$	$p = .006$**	$p = .015$*	$p = .17$	$p = .008$**	$p = .017$*

Notes: BPVS-3 = British Picture Vocabulary Scale-3rd Edition; PLSI = Pragmatic Language Skills Inventory; PIS = Personal Interaction Skills; SIS = Social Interaction Skills; CI = Classroom Interactions; ToMI-2 = Theory of Mind Inventory-2; Early = Early Subtest; Basic = Basic Subtest; Advanced = Advanced Subtest; MS Terms = Mental State Terms Subtest; Emo. Recog. = Emotion Recognition Subtest; Pragmatics = Pragmatics Subtest.
*$p < .05$; **$p < .01$; [a]$p < .10$ (significant with alpha = .10).

advantage of the ToMI-2, which is a developmentally sequenced broadband measure of ToM, is that it cannot only identify ToM strength and challenge areas, but it generates an analysis for each individual that is rooted in an empirically driven understanding of ToM development (see Theoryofmindinventory.com for sample reports that offer recommendations for the most developmentally appropriate treatment goals based on individual results). As such, use of the ToMI-2 facilitates a descriptive-developmental approach to intervention which is a common standard in clinical sciences (Paul 2007) including communication sciences and disorders, clinical psychology, and special education. A descriptive-developmental approach is one in which the typical developmental sequence of skills in a particular domain is described in detail. The developmental sequence is then compared to the results of a comprehensive assessment for a given individual to determine the earliest skills that are in need of support. This is important because early skills tend to be more readily supported and are foundational to later developing skills.

This approach is also considered best practice because it is designed to scaffold learning in the zone of proximal development (ZPD). Vygotsky (1978) described the ZPD as 'the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance, or in collaboration with more capable peers' (p. 86). Indeed, professionals who have over- or under-estimated the developmental level of a student or client will be well-aware of the benefits of working in the ZPD to not only advance learning but to also decrease frustration and motivate engagement. Using a descriptive-developmental approach, the ToMI-2 can support a clinician's understanding of ToM progression and identification of treatment priorities to facilitate immediate and long-term goal setting.

A related resource, known as the ToM Atlas is also available at no cost to help professionals understand and communicate about the nature and development of distinct ToM areas. Free access to the ToM Atlas is available after registration at theoryofmindinventory.com. The ToM Atlas is an ever-expanding evidence-based online resource designed to map, define, and describe the development of ToM in typical development, autism, attention-deficit disorders, and those with hearing loss. Ultimately, our hope is that clinicians will find the ToMI-2 useful for identifying developmentally appropriate treatment goals and will access the ToMA for understanding the nature and development of each ToM domain in greater detail. This grounding in ToM development and empiricism presents a practical starting point for developing interventions (e.g. interactive reading programmes, thought-bubble exercises; Wellman and Peterson 2013, Stanzione and Schick 2014, Beazley and Chilton 2015) that have been shown to be effective and which are feasible in real-world classroom practice.

As this pilot study is the first to formally evaluate the utility of the ToMI-2 for a sample of CHL, our second purpose was to gather preliminary data to inform the potential for the tool's criterion-related validity for this population. As expected, ToMI-2 composite scores correlated with scores on an index of pragmatic language as well as hearing vocabulary. In addition, the ToMI-2 composite score correlated strongly and negatively with both age of amplification and age of implantation. Moreover, three ToMI-2 subtests (the Advanced subscale, and the Mental State Term Comprehension and Pragmatics subscales) correlated with age of implantation. Although correlations with age of amplification were approaching significance and would likely be significant with a larger sample, these correlations did not achieve statistical significance in this pilot study.

These data bode well for the ToMI-2 as a construct valid indicator of ToM in CHL although one might wonder why the ToMI-2 composite, Advanced, Mental State Term, and Pragmatic subscales correlations with age of implantation achieved significance while the correlations for the Early, Basic, and Emotion Recognition subscales did not. We suspect that two effects are operating. First, we expect that effects for all ToMI-2 subscales will achieve significance for all criteria comparisons in the future once appropriately powered statistical tests are conducted (a sample size of 50 should be adequate). Second, our sample was comparable in age (ages 5 years 2 mo — 11 years 1 mo, $M = 8.5$ years), language ability (which was diverse), and hearing history to most other studies examining ToM in CHL. This gains importance in light of the fact that the age of 8-years is when typically developing children would be undergoing rapid development in Advanced ToM skills as assessed on the ToMI-2. Moreover, older children have likely had sufficient time and social experience to overcome challenges in early ToM development. As a result, a restriction of range in those early ToM abilities may have obscured effects that would be evident in a sample of younger CHL. This interpretation is consistent with the conversation hypothesis of ToM challenges in CHL although future research is needed to formally evaluate the links we propose. Other directions for future validation of the ToMI-2 for use with CHL involves tests of social validity (from caregivers and professionals providing services to CHL) to evaluate the measures perceived effectiveness and practicality. In addition, several other tests of criterion-related validity are needed. Most notably, correlations between the ToMI-2 and measures of direct child performance using the traditional ToM batteries are sorely needed.

The current study provides tentative evidence for the ToMI-2 as a useful measure of ToM competence that can be used in concert with or in lieu of traditional ToM tasks. Besides simply not being readily available, traditional ToM tasks can be difficult to administer especially to CHL who vary greatly in their hearing and instructional needs. Even when ToM task batteries are used, scores are most appropriately construed as ordinal in nature, and concerns arise over test-practice effects making the measures questionable when used as a repeated measure. By contrast, the ToMI-2 is quick and easy to administer, yields interval data, may be used as a repeated measure, and may provide finer levels of discrimination across a broader range of development and skill levels. It may also be particularly useful as a socially valid measure to be used in the context of intervention studies targeting ToM and it is considered a family centered assessment that can be used to begin a conversation with caregivers towards a better understanding of the child's strengths and challenges as well as the families' priorities for intervention. Given the good psychometric properties demonstrated by the ToMI-2 thus far, this study provides support for further examination and development of the ToMI-2 as a research and clinical tool to support CHL.

Notes

1. Another popular false belief task is known as the unexpected contents task in which a child is shown surprising contents of a container (e.g., candy in a pencil box) and is asked what a naïve observer (one who has not looked inside the container) will predict is in the box.
2. The procedures used in the first study on CHL (Peterson et al. 2005) slightly adapted the tasks borrowed from Wellman and Lui (2004).

Conflict of interest

Tiffany Hutchins is the lead author and developer of the Theory of Mind Inventory and co-owner of Theory of Mind Inventory, LLC.

References

Astington, J.W. and Baird, J.A., 2005. Introduction: why language matters. *In*: J.W. Astington and J.A. Baird, eds. *Why language matters for theory of mind*. New York: Oxford University Press, 3–25.

Beazley, S. and Chilton, H., 2015. The voice of the practioner: Sharing fiction books to support the understanding of theory of mind in deaf children. *Deafness and Education International*, 17 (4), 231–240.

British Educational Research Association. 2011. *Ethical guidelines for educational research*. London: BERA.

Cornes, A., and Brown, M., 2012. Mental health of Australian deaf adolescents: an investigation using an Auslan version of the strengths and difficulties questionnaire. *Deafness and Education International*, 14 (3), 161–175.

Crais, E., 1993. Families and professionals as collaborators in assessment. *Topics in Language Disorders*, 14, 29–40.

Dunn, L.M. and Dunn, L.M., 2007. *Examiner's manual for the peabody picture vocabulary test*, 4th ed. Circle Pines, MN: American Guidance Service.

Dunn, L., Dunn, L., Sewell, J., Styles, B., Brzyska, B., Shamsan, Y., and Burge, B., 2009. *The British picture vocabulary scale*. 3rd ed. London: GL Assessment.

Falkman, K., Roos, C., and Hjelmquist, E., 2007. Mentalizing skills of non-native, early signers: A longitudinal perspective. *European Journal of Developmental Psychology*, 4 (2), 178–197.

Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, J., and Bates, E., 2007. *The MacArthur-Bates communicative development inventories*. Baltimore, MD: Brookes.

Figueras-Costa, B. and Harris, P., 2001. Theory of mind development in deaf children: a nonverbal test of false-belief understanding. *Journal of Deaf Studies and Deaf Education*, 6, 92–102.

Gilliam, J. and Miller, L., 2006. *The pragmatic language skills inventory*. Austin, TX: Pro-Ed.

Gonzalez, A., Quintana, I., Barahas, C., and Linero, M., 2007. The role of age and oral lexical competence in false belief understanding by children and adolescents with hearing loss. *The Volta Review*, 107 (2), 123–139.

Gray, C., Hosie, J., Russell, P., Scott, C., and Hunter, N., 2007. Attribution of emotions to story characters by severely and profoundly deaf children. *Journal of Developmental and Physical Disabilities*, 19 (2), 145–159.

Greenslade, K. and Coggins, T., 2016. Brief report: an independent replication and extension of psychometric evidence supporting the theory of mind inventory. *Journal of Autism and Developmental Disorders*, 46 (8), 2785–2790.

Hao, J. Su, Y., & Chan, R., (2010). Do deaf adults with limited language have advanced theory of mind? *Research in Developmental Disabilities*, 31, 1491–1501.

Hutchins, T. and Prelock, P., 2016. *Technical manual for the theory of mind inventory-2*. Unpublished Copyrighted Manuscript. Available at: theoryofmindinventory.com.

Hutchins, T., Prelock, P., and Bonazinga, L., 2012. Psychometric evaluation of the Theory of Mind Inventory (ToMI): A study of typically developing children and children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 42, 327–341.

Hutchins, T., Prelock, P., Bonazinga, L., and Taylor, R., 2008. Beyond false beliefs: the development and psychometric evaluation of the Perceptions of Children's Theory of Mind Measure — Experimental version (PCToMM-E). *Journal of Autism and Developmental Disorders*, 38, 142–155.

Klin, A., 2000. Attributing social meaning to ambiguous visual stimuli in higher-functioning autism and Asperger Syndrome: the social attribution task. *Journal of Child Psychology and Psychiatry*, 7, 831–846.

Lerner, M., Hutchins, T., and Prelock, P., 2011. Brief report: preliminary evaluation of the theory of mind inventory and its relationship to measures of social skills. *Journal of Autism and Developmental Disorders*, 41, 512–517.

Macauley, C. and Ford, R., 2006. Language and theory-of-mind development in prelingually deafened children with cochlear implants: a preliminary study. *Cochlear Implants International*, 7 (1), 1–14.

McCauley, R., 2001. *Assessment of language disorders in children*. Mahwah, NJ: Lawrence Erlbaum.

Most, T., Shina-August, E., and Meilijson, S., 2010. Pragmatic abilities of children with hearing loss using cochlear implants or hearing aids compared to hearing children. *Journal of Deaf Studies and Deaf Education*, 15 (4), 422–437.

Netten, A., Rieffe, C., Theunissen, S., Soede, W., Dirks, E., Briaire, J., and Frijns, J., 2015. Low empathy in deaf and hard of hearing (pre)adolescents compared to normal hearing controls. *PLoS ONE*, 10 (4), 1–15.

Paul, R., 2007. *Language disorders from infancy through adolescence*. 3rd ed. Oxford, England: Mosby Elsevier.

Peters, K., Remmel, E., and Richards, D., 2009. Language, mental state vocabulary, and false belief understanding in children with cochlear implants. *Language Speech and Hearing Services in Schools*, 40 (3), 245–255.

Peterson, C., and Siegal, M., 1995. Deafness, conversation, and theory of mind. *Journal of Child Psychology and Psychiatry*, 36 (3), 459–474.

Peterson, C., and Slaughter, V.P., 2006. Telling the story of theory of mind: Deaf and hearing children's narratives and mental state understanding. *British Journal of Developmental Psychology*, 24, 151–179.

Peterson, C. and Wellman, H., 2009. From fancy to reason: Scaling deaf and hearing children's understanding of theory of mind and pretence. *British Journal of Developmental Psychology*, 27, 297–310.

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

Peterson, C.C., Wellman, H.M., and Slaughter, V., 2012. The mind behind the message: advancing theory of mind scales for typically developing children and those with deafness, autism, or Asperger syndrome. *Child Development*, 83, 469–485.

Pyers, J. and de Villiers, P., 2013. Theory of mind in deaf children: illuminating the relative roles of language and executive functioning in the development of social cognition. *In*: Baron-Cohen, S., Tager-Flusberg, H., and Lombardo, M., eds. *Understanding other minds: perspectives from developmental social neuroscience*. New York: Oxford University Press, 345–363.

Pyers, J. and Senghas, A., 2009. Language promotes false-belief understanding. *Psychological Sciences*, 20 (7), 805–811.

Remmel, E., & Peters, K., 2008. Theory of mind and language in children with cochlear implants. *Journal of Deaf Studies and Deaf Education*, 14 (2), 218–236.

de Rosnay, M., Fink, E., Begeer, S., Slaughter, V., and Peterson, C., 2014. Talking theory of mind talk: young school-aged children's everyday conversation and understanding of mind and emotion. *Journal of Child Language*, 41, 1179–1193.

Schick, B., de Villiers, P., de Villiers, J., and Hoffmeister, R., 2007. Language and theory of mind: a study of deaf children. *Child Development*, 78 (2), 376–396.

Stanzione, C. and Schick, B., 2014. Environmental language factors in theory of mind development: evidence from children who are deaf/heard of hearing or who have specific language impairment. *Topics in Language Disorders*, 34, 296–312.

Sundqvist, A., Lyxell, B., Jonsson, R., and Heimann, M., 2014. Understanding minds: early cochlear implantation and the development of theory of mind in children with profound hearing impairment. *International Journal of Pediatric Otorhinolaryngology*, 78, 538–544.

de Villiers, J., and de Villiers, P., 2014. The role of language in theory of mind development. *Topics in Language Disorders*, 34 (4), 313–328.

de Villiers, P., and de Villiers, J., 2000. Linguistic determinism and the understanding of false beliefs. *In*: P. Mitchell and K. Riggs, eds. *Children's reasoning and the mind*. Hove, UK: Psychology Press, 191–228.

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

Wellman, H. and Lui, D., 2004. Scaling of theory-of-mind tasks. *Child Development*, 75 (2), 523–541.

Wellman, H. and Peterson, C., 2013. Deafness, thought bubbles, and theory-of-mind development. *Developmental Psychology*, 49 (12), 2357–2367.

Wimmer, H. and Perner, J., 1983. Beliefs and beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.

Woolfe, T., Want, S., and Siegal, M., 2002. Signposts to development: theory of mind in deaf children. *Child Development*, 73 (3), 768–778.