Cognitive Development 68 (2023) 101366

Contents lists available at ScienceDirect

Cognitive Development

journal homepage: www.elsevier.com/locate/cogdev

The Early Independent Problem Solving Survey (EIPSS): Its psychometric properties in children aged 12–47 months

Elena Hoicka [a,*], Stephanie Powell [b], Sarah E. Rose [c], Eva Reindl [d,e], Claudio Tennie [f]

[a] School of Education, University of Bristol, Bristol, UK
[b] Department of Psychology, University of Sheffield, Sheffield, UK
[c] School of Health, Science and Wellbeing, Staffordshire University, UK
[d] Department of Anthropology, Durham University, UK
[e] School of Psychology and Neuroscience, University of St Andrews, UK
[f] Institut für Ur, und Frühgeschichte und Archäologie des Mittelalters, Eberhard Karls Universität Tübingen, Germany

A R T I C L E I N F O

Keywords:
Problem solving
Creativity
Innovation
Infancy
Todder
Preschool

A B S T R A C T

Independent problem solving (IPS) involves solving problems alone; with motivation and persistence; without watching others; or requesting or receiving help. The Early Independent Problem Solving Survey (EIPSS) was developed for 12- to 47-month-olds. Study 1 ($N = 272$) found good internal reliability and a 2-factor structure: Repetitive (repeatedly solvable problems, e.g., jigsaws) and Novel IPS (one-off problems, e.g., reaching out-of-reach toys). Study 2 ($N = 567$) confirmed good internal reliability and the 2-factor structure. Study 3 ($N = 85$) found a positive correlation between a divergent thinking lab measure and Novel IPS. Study 4 found good 6-month-longitudinal stability ($N = 110$) for the EIPSS and its subscales; and good agreement between parents ($N = 32$) for the Repetitive subscale. Study 5 (all data combined) demonstrated no item functioning differences across demographic variables. Differences for child age, child gender, parent age, and parent education were found for the EIPSS and subscales.

1. Introduction

Independent problem solving (IPS) is defined as a way to find a solution to a problem alone, without watching others or asking for help, while being motivated and persistent (Beck et al., 2016; Day & Burns, 2011; Redding et al., 1988; Sigman et al., 1987; Thompson, 1999; Thompson et al., 2012; Thompson & Moore, 2000; Vlachou & Farrell, 2000; Yarrow et al., 1982). While research often focuses on young children's ability to learn from others (i.e., social learning, such as imitation), baseline conditions of several social learning experiments show children are also capable of solving problems on their own (Bechtel et al., 2013; Call et al., 2005; Dean et al., 2012; Fagard et al., 2014; Flynn et al., 2016; Rat-Fischer et al., 2012; Reindl et al., 2020; Speidel et al., 2021; Subiaul et al., 2015; Tennie et al., 2010). However, it is unclear to what extent and how young children independently solve problems in real life. Our goal is to create and evaluate a new parent-report measure of early IPS: the Early Independent Problem Solving Survey (EIPSS). This tool will allow researchers to begin to understand how IPS develops in early life.

Children under 1 year solve both physical and social problems independently in their natural environment (Keen, 2011). For example, 4- to 18-month-olds solved 89 % of obstacles by manipulating objects, using tools or force (physical); or by complying with

* Correspondence to: School of Education, University of Bristol, Bristol BS8 1JA, UK.
E-mail address: elena.hoicka@bristol.ac.uk (E. Hoicka).

https://doi.org/10.1016/j.cogdev.2023.101366
Received 16 March 2022; Received in revised form 7 August 2023; Accepted 10 August 2023
Available online 7 September 2023
0885-2014/

others, refusing others, or trying to get others to do something (social) (Henderson & Dias, 1987). Experiments show 7-month-olds retrieve out-of-reach objects, by e.g., removing barriers and pulling supports (Babik et al., 2019; Chen et al., 1997; Willatts, 1984, 1999).

One-year-olds determine how new toys work (Yarrow et al., 1982), complete simple puzzles (Bono & Stifter, 2003; Redding et al., 1988; Yarrow et al., 1982), build towers (Bono & Stifter, 2003; Chen et al., 2010), get through a barrier to retrieve a prize (Fidler et al., 2005; Yarrow et al., 1982), and continue to reach out-of-reach objects (Fagard et al., 2014; Rat-Fischer et al., 2012, 2014). Additionally, 13-month-olds spent 60 % of their time working on their tasks, a marker of persistence, and their persistence correlated with task success (Yarrow et al., 1982). Furthermore, 1-year-olds' motor actions suggest that when building towers, they plan ahead (Chen et al., 2010).

Two-year-olds continue to complete puzzles, using trial and error, shape matching, and color matching as strategies (Levine et al., 2012; Montford & Readdick, 2008). They can solve puzzle boxes alone (Call et al., 2005; Hopper et al., 2020; Seed & Call, 2014; Sigman et al., 1987; Speidel et al., 2021; Tennie et al., 2010). Finally, 2-year-olds begin independently choosing the correct tools to obtain out of reach objects (Breyel & Pauen, 2021; Chen & Siegler, 2013; Gardiner et al., 2012; Neldner et al., 2020; Reindl et al., 2016; Seed & Call, 2014). Additionally, children are more successful at removing out-of-reach objects with their fingers than tools, suggesting tool use is more difficult for 2-year-olds, possibly due to e.g., an overload of working memory capacity (Seed & Call, 2014).

Three-year-olds continue to solve puzzles (Rose et al., 2021), and now do so by matching different patterns (Pepler & Ross, 1981). Furthermore, their puzzle solving speed relates to their mastery-orientation, that is, they see it as an opportunity to increase their competence, rather than showing others what they can do (Day & Burns, 2011). They continue to retrieve out-of-reach objects with appropriate tools (Gredlein & Bjorklund, 2005), and solve puzzle boxes (Dean et al., 2012; Flynn et al., 2016; Reindl et al., 2020; Subiaul et al., 2015; Vlachou & Farrell, 2000). They also solve weight puzzles (Metz, 1993), independently solve the Tower of Hanoi (Tarasuik et al., 2017) and solve social problems, e.g., convincing someone to play with them or give them an object (Krasnor & Rubin, 1983; Neel et al., 1990). Most 3-year-olds fail to innovate tools to problem solve, such as re-shaping or joining tools (Beck et al., 2016; Breyel & Pauen, 2021; Gönül et al., 2018).

Overall, research suggests children independently solve problems in a variety of ways from infancy. However, little is known about how they independently solve problems in their real lives. One way to capture this is to create a parent-report measure of IPS. Indeed, the Ages and Stages Questionnaire (ASQ) Version 3 for 4- to 60-month-olds includes a Problem Solving subscale (Squires & Bricker, 2009). However, it is not specifically an *independent* problem solving subscale. While one to two items across each age range capture IPS, e.g., trying to get a cheerio out of a clear bottle, other aspects capture problem solving more broadly, e.g., gaze following, imitation, looking at objects, vocabulary (e.g., colors and numbers), and pretending. Therefore, an IPS parent-report measure is still needed. The Problem Solving Questionnaire (PSQ) (Camp et al., 2016) examines real life problem solving in typical and atypical development, asking about children's abilities in real life tasks including dressing, brushing teeth, making sandwiches, and making phone calls; and novel tasks including finding lost possessions, packing bags, and putting away items in a wardrobe or chest of drawers. For each item, they ask if children recognize the problem; know the goal of the problem; have ideas about how to solve the problem; keep track of what they are doing; stay focused on the task; use existing knowledge to help them; apply a strategy previously learnt in another situation; and stop when tasks are finished. They also ask whether, if something goes wrong, children: ask for help, lose focus, stop persevering, and keep going until they reach a solution. Therefore, implicitly, it appears the PSQ has an *independent* problem solving focus. However, it was designed for adolescents with atypical development, and children from around 7 or 8 years with typical development. Therefore, the PSQ is likely too advanced for children under 4 years.

We therefore decided to create an ecologically valid parent-report measure of early IPS: the EIPSS. Our primary goal was to create a tool allowing researchers to efficiently learn more about early IPS and related areas of development. For instance, one emerging area of research, first put forward by Vygotsky (1978), considers the relationship between social learning and innovation (Bonawitz et al., 2011; Carr et al., 2015; Hoicka et al., 2018; Legare & Nielsen, 2015; Rawlings et al., 2017; Rawlings et al., 2021). Researchers already running an imitation lab task could easily learn more about this relationship by giving parents a short IPS survey; or parents could complete an IPS survey alongside another measure capturing early social cognition, such as the Early Social Cognition Inventory or the Children's Social Understanding Scale (Hoicka et al., 2021; Tahiroglu et al., 2014). Another goal was to create an efficient measure of IPS to act as a control measure for experiments involving IPS, e.g., tool use, or innovation, which could give experiments more rigor. For example, parents could complete an IPS survey before children participate in a between-participants experiment involving IPS, and the IPS survey could act as a baseline measure to control for pre-existing differences between participant groups. Our goal was to cover a wide age range so both cross-sectional and longitudinal research could use the same survey to tap into developmental change. Another goal was creating a survey useful across English-speaking countries as the research discussed gives evidence from a variety of countries, including Australia, Canada, the UK, and the USA. Creating a survey useful across English-speaking countries allows for future research across various English-speaking countries, and research collaborations across those countries.

A final goal was to use the EIPSS to explore demographic differences. First, we expected IPS to increase with age, in line with past research (Babik et al., 2019; Barrett et al., 2007; Breyel & Pauen, 2021; Gardiner et al., 2012; Gönül et al., 2018; Hopper et al., 2020; Montford & Readdick, 2008; Neldner et al., 2019; Rat-Fischer et al., 2014; Rat-Fischer et al., 2012; Reindl et al., 2016; Seed & Call, 2014; Tarasuik et al., 2017; Willatts, 1984, 1999). Second, we sought to explore gender differences. Many early IPS studies found no gender differences (Chen et al., 2010; Chen et al., 1997; Chen & Siegler, 2013; Day & Burns, 2011; Fidler et al., 2005; Gardiner et al., 2012; Gönül et al., 2018; Levine et al., 2012; Neldner et al., 2019; Pepler & Ross, 1981; Reindl et al., 2016; Thompson, 1999). However, some studies found boys outperform girls (Barrett et al., 2007; Bates et al., 1980), although another study found the reverse (Thompson et al., 2012). Furthermore, older (5- to 6-year-old) boys were more likely to innovate tools than girls in Turkey, although not in New Zealand (Gönül et al., 2019). Third, we sought to explore socio-economic status (SES) differences. Higher parental income, but not parental education, was linked to increased IPS in young children (Day & Burns, 2011; Levine et al., 2012). In older children, higher SES as a combination of parents' income, education, and job type, correlated with higher IPS (Burns et al., 1987); children from higher SES areas outperformed children from lower SES areas (Cox, 1985); and increased parental education sometimes predicted IPS, but not always (Greiff et al., 2013; Hacatrjana, 2022). Fourth, we sought to explore cultural differences, including country, specifically, Australia, Canada, the United Kingdom, and the United States; as well as by language status (mono- versus multilingual). While we are unaware of research comparing young children's IPS in these particular countries, previous research found IPS differences across other countries. Australian 2- to 5-year-olds outperformed South African children, from villages, townships, and Bushmen, on IPS tasks (Neldner et al., 2019; Neldner et al., 2020). Children (5–6 years) in New Zealand outperformed Turkish children on an IPS task (Gönül et al., 2019). In contrast, remote Indigenous and Western urban Australian 3- to 5-year-olds showed no differences (Neldner et al.,

Table 1
EIPPS Item number; question; source; Pearson correlations between the final items and entire survey (Study 1); factor loadings (F1, F2) for the EFA (Study 1), CFA for Study 2 (in brackets), and CFA for Study 3 (in brackets, in Italics). Parents answered on a 4-point Likert scale from 1 (Strongly Disagree) to 4 (Strongly Agree), or Not Applicable for each item. Items were not included if at least 20 % of parents answered Not Applicable for an item, or if items did not correlate with the EIPSS mean score above r = 0.3 (see Study 1, Results).

Item	Question	Source	*r*	F1	F2
1	My child makes difficult structures with blocks/ megablocks on his/her own	(Bono & Stifter, 2003; Chen et al., 2010; Gredlein & Bjorklund, 2005)	.59	**.42** (.58) (*.22*)	.14
5	My child sorts objects in logical ways on his/her own, e.g., cutlery next to plates; toy bed in dollhouse	(Camp et al., 2016)	.72	**.52** (.60) (*.51*)	.24
6	My child puts away objects correctly on his/her own	(Camp et al., 2016; Squires & Bricker, 2009)	.63	**.41** (.47) (*.13*)	.22
8[R]	My child needs help to sort objects by, e.g., shape, colour	(Camp et al., 2016; Metz, 1993; Montford & Readdick, 2008; Pepler & Ross, 1981; Vlachou & Farrell, 2000)	.66	**.89** (.73) (*.41*)	-.16
10[R]	My child needs help completing simple puzzles	(Bono & Stifter, 2003; Day & Burns, 2011; Levine et al., 2012; Montford & Readdick, 2008; Pepler & Ross, 1981; Redding et al., 1988; Rose et al., 2021; Tarasuik et al., 2017; Thompson, 1999; Thompson et al., 2012; Yarrow et al., 1982)	.69	**.74** (.78) (*.67*)	.01
2[R]	My child needs help from others when figuring out how new toys work	(Dean et al., 2012; Flynn et al., 2016; Reindl et al., 2020; Speidel et al., 2021; Subiaul et al., 2015; Tennie et al., 2010; Vlachou & Farrell, 2000; Yarrow et al., 1982)	.65	.20	**.52** (.69) (*.43*)
3[R]	When it is physically difficult to reach a toy, my child will not try to figure out how to reach it	(Babik et al., 2019; Breyel & Pauen, 2021; Chen et al., 1997; Chen & Siegler, 2013; Fagard et al., 2014; Fidler et al., 2005; Gardiner et al., 2012; Gredlein & Bjorklund, 2005; Hopper et al., 2020; Neldner et al., 2019; Neldner et al., 2020; Rat-Fischer et al., 2012, 2014; Reindl et al., 2016; Seed & Call, 2014; Squires & Bricker, 2009; Tennie et al., 2010; Willatts, 1984, 1999; Yarrow et al., 1982)	.56	.12	**.43** (.36) (*.59*)
4[R]	My child does not like to take things apart and put them back together	(Call et al., 2005; Hopper et al., 2020; Neldner et al., 2019; Subiaul et al., 2015)	.68	.11	**.64** (.43) (*.65*)
7	My child looks at the mechanisms of how things work, e.g., how wheels are connected to toy cars	(Beck et al., 2016; Breyel & Pauen, 2021; Chen & Siegler, 2013; Dean et al., 2012; Fidler et al., 2005; Gardiner et al., 2012; Gönül et al., 2018; Hopper et al., 2020; Metz, 1993; Reindl et al., 2020; Seed & Call, 2014; Subiaul et al., 2015)	.56	-.20	**.80** (.51) (*.40*)
9	If a toy breaks, my child tries to fix it before asking for help	(Sigman et al., 1987; Thompson, 1999; Thompson et al., 2012; Thompson & Moore, 2000)	.61	.10	**.52** (.62) (*.57*)
	Items not included in EIPSS:				
11	When a toy breaks, my child uses past experiences to try to fix it	(Chen & Siegler, 2013)			
12	My child resolves conflicts with other children on his/ her own	(Krasnor & Rubin, 1983; Neel et al., 1990)			
13[R]	My child needs help to come up with good plans, e.g., figuring out that s/he needs an umbrella if it's raining	(Camp et al., 2016)			
14[R]	If my child draws a picture wrong, s/he gives up instead of starting over again	(Sigman et al., 1987; Vlachou & Farrell, 2000; Yarrow et al., 1982)			
15	When my child has a question, s/he tries to figure out the answer him/herself before going to someone else	(Sigman et al., 1987; Thompson, 1999; Thompson et al., 2012; Thompson & Moore, 2000)			
16[R]	If my child does not get what s/he wants from me, s/ he does not ask someone else, e.g., other parent, grandparent	(Sigman et al., 1987; Vlachou & Farrell, 2000; Yarrow et al., 1982)			
17[R]	When doing crafts, my child needs help to fix things if things go wrong, e.g., if crayon too short, needs help to pull off paper	(Camp et al., 2016; Squires & Bricker, 2009)			

[R] Item is reverse-coded.

Table 2
Participant information.

	Study 1	Study 2	Study 3
N	272	567	85
Children's Age:			
Mean (months; days)	28;5	26;24	36;13
Range	12;4–47;26	12;0–47;24	24;20–47;25
SD	10;10	9;28	7;4
Children's Gender:			
Female	134	289	41
Male	138	276	44
Not reported	0	2	0
Children's Ethnicity:			
Arab	0	1	0
Black	5	5	0
East Asian	6	1	0
Hispanic	1	0	0
Pacific Islander	0	1	0
South Asian	4	5	2
West Indian	0	1	0
White	198	507	78
Of Mixed Ethnicity	1	21	5
Other (not specified)	11	19	0
Not reported	46	6	0
Country:			
Australia	20	5	0
Canada	5	11	0
United Kingdom	183	462	85
United States of America	35	62	0
Other	24	21	0
Not reported	8	6	0
Child's Language			
English only	139	459	41
English and another language(s)	41	83	8
One other language only	0	12	0
Not reported	92	13	36
Sibling(s)			
No	114	292	15
Yes	68	263	34
Not reported	90	12	36
Childcare			
Mean (hours)	*N/A*	17.28	17.07
Range		0–75	0–50
SD		14.21	11.51
Not reported	272	53	36
Parents' Age			
Mean (years)	33.84	33.96	35.40
Range	18–48	18–67	24–45
SD	5.20	5.17	4.86
Not reported	88	14	38
Parents' Gender			
Female	183	511	48
Male	4	40	1
Not reported	85	16	36
Parents' Ethnicity:			
Black	5	7	0
East Asian	6	8	0
South Asian	2	4	1
White	163	511	44
Of Mixed Ethnicity	1	7	1
Other (not specified)	8	15	0
Not reported	87	15	36
Parents' Education			
Secondary School	36	68	15
Community College	8	26	0
Undergraduate Degree	72	207	38
Postgraduate Degree	108	254	31
Not reported	48	12	1
Household Income			
Australia: N	11	5	*NA*
Mean	$129,091AUD	$120,000	

(*continued on next page*)

Table 2 (*continued*)

	Study 1	Study 2	Study 3
Range	$40,000 – $200,000	$50,000-$200,000	
SD	$54,627	$65,955	
Canada: N	*NA*	9	*NA*
Mean		$129,444CAD	
Range		$50,000-$200,000	
SD		$53,411	
United Kingdom: N	63	317	46
Mean	£ 54,238GBP	£ 58,400	£ 49,949
Range	£ 5000–£ 120,000	£ 6000–£ 750,000	£ 17,210–£ 112,000
SD	£ 26,639	£ 49,006	£ 25,187
United States of America: N	31	53	*NA*
Mean	$152,839USD	$125,453	
Range	$24,000–$500,000	$20,000–$250,000	
SD	$89,798	$59,119	
Recruited			
babylovesscience.com	214	441	0
Sheffield Cognitive Development Lab	41	126	85
The Psychology Children's Lab, Staffordshire University	43	0	0

2017). We are also unaware of research linking multilingualism to IPS in this age group. However, research on older children and adults suggests mixed results (Bialystok & Majumder, 1998; Cushen & Wiley, 2011; Stafford, 1968). Sixth, we sought to explore social aspects of IPS, including having siblings and the number of hours spent in childcare. There is little research in this area, however one study found no difference in IPS in children with more versus fewer siblings (Cicirelli, 1976). Finally, we sought to examine how parental demographics affect reporting – specifically, parent age and gender. For instance, mothers and fathers sometimes report differently on young children, but often do not, in areas such as social cognition, socio-emotional development, problem eating, and abnormal social behaviors (Adamson & Blight, 2014; Alakortes et al., 2015; Davé et al., 2008; Gluck et al., 2021).

The EIPSS has important real-world implications. First, early years curricula emphasize problem solving in several countries, including Australia, Canada, the UK, and the USA (Australian Government Department of Education and Training, 2017; Best Start Expert Panel on Early Learning, 2007; Department for Education, 2017; Ohio Department of Education, 2012). Having a tool to better understand early IPS could allow us to determine if, e.g., children from lower SES backgrounds need more support in developing IPS skills. Second, this research could be useful to parents, early years educators, and medical staff, such as health visitors and paediatricians, to know which IPS behaviors to expect, and, with future development of the EIPSS, to potentially serve as a marker of development difference.

This paper sought to develop a parent report measure of IPS from 12 to 47 months. Study 1 used Exploratory Factor Analysis (EFA, $N = 272$) to determine the EIPSS's internal reliability and factor structure. Study 2 used Confirmatory Factor Analysis (CFA, $N = 574$) to replicate the internal reliability and factor structure. Study 3 ($N = 85$) examined convergent validity by comparing the EIPSS to children's performance on an IPS task (Great Ape Tool Test Battery, GATTeB) and a divergent thinking task (Unusual Box Test, UBT) in the lab (Bijvoet-van den Berg & Hoicka, 2014; Hoicka et al., 2016; Reindl et al., 2016). Study 4 used a subsample of participants from Studies 1 and 2 to measure inter-observer reliability between parents ($N = 32$), and 6-month longitudinal stability ($N = 110$). Study 5 examined data from the first three studies together ($N = 83$–924) to examine similarities and differences in item functioning, and overall scores across demographic variables (e.g., child gender, parent education, household income).

2. Study 1: survey construction

The initial survey was designed to capture early IPS from 4 months, or even earlier, based on Henderson and Dias (1987) observing IPS as early as 4 months. However, pilot data for children under 12 months suggested too many questions were answered "not applicable", therefore the survey was limited to children between 12 and 47 months. The upper age limit was chosen as the focus of the EIPSS was IPS before school age, and British children, the main sample, begin school from 48 months. PsycInfo was searched for terms including "problem solv* ," "convergent thinking," "creativ* ," and "innovat* " alongside terms such as "preschool* ," "toddler* ," and "infan* ." and items were created based on relevant papers, e.g., items relating to puzzles, construction, retrieving out-of-reach objects, determining how new toys work, resolving conflicts with children, all within real life concepts appropriate to young children, e.g., toys, eating, crafts (see Table 1). We also took inspiration from the ASQ and PSQ (Camp et al., 2016; Squires & Bricker, 2009). In writing the items, problem solving should be *independent*, and so terms such as "alone," "needs help," "try," "gives up," and "likes to," were embedded to get at the concept of IPS being individual, motivated, persistent, and involving mastery (Beck et al., 2016; Day & Burns, 2011; Redding et al., 1988; Sigman et al., 1987; Thompson, 1999; Thompson et al., 2012; Thompson & Moore, 2000; Vlachou & Farrell, 2000; Yarrow et al., 1982). This led to 17 items for the initial EIPSS (see Table 1), which used a 4-point Likert scale, from "Strongly Agree" to "Strongly Disagree", with some items reverse scored. We chose a 4-point scale to avoid parents staying neutral. A couple of parents gave written feedback about the EIPSS after completing it to ensure questions made sense to them.

2.1. Method

2.1.1. Participants

See Appendix A for power analysis. Surveys were obtained for 272 children. Data were collected from February 2015 to October 2016. See Table 2 for participant information. Income statistics of samples with fewer than five participants per country were not reported. Participants were recruited through Facebook advertising, targeting parents of children 1–3 years in English-speaking countries, posts on lab and parenting Facebook pages, social media, press releases, Bounty packs in Sheffield, UK, and letters to parents sent through nurseries in Staffordshire. Ethical approval was obtained from the Psychology Department at the University of Sheffield for the projects, “Using parent reports to learn about early humour, pretending, deception, creativity, social cognition, actions, and language”, Reference Number 003095, and “Young children’s technology use & creativity”, Reference Number 012752, and from the School of Health, Science and Wellbeing, Staffordshire University for the project "The immediate effect of pace television on children’s creativity". Parents who completed the survey on babylovesscience.com ticked boxes online to indicate consent. Parents who completed the survey in the lab signed paper consent forms. Participants from the Sheffield Cognitive Development lab received a book for participating in the survey alongside an unrelated lab study. Participants from babylovesscience.com received no reward, unless participants repeated the survey 6 months later, or the child’s other parent also completed the survey (see Study 4). Participants from The Psychology Children’s Lab, Staffordshire University, received £10 and a t-shirt for participating in the survey alongside an unrelated lab study.

2.1.2. Measure

2.1.2.1. Early Independent Problem Solving Survey (EIPSS; preliminary version). Seventeen questions asked how keen children are to do IPS in real life, e.g., “If a toy breaks, my child tries to fix it before asking for help” (see Table 1). The beginning of the survey stated:

> “For each statement, think back to specific examples in the last month, and choose whether you: Strongly Agree; Agree; Disagree; Strongly Disagree. However, sometimes you will not be able to think of a specific time in the last month to evaluate the statement. For instance, if we ask about how your child responded to getting a present, and your child didn’t get a present in the last month, skip the question or choose NA (not applicable).”

Responses were averaged to give an EIPSS score. Items for which parents answered “NA” were not included.

2.2. Results

EIPSS items for Sample 1 ($N = 272$) were removed for which more than 20 % of participants answered “NA” as these items may be too developmentally advanced for younger participants or may not be popular with or encouraged by parents. Items 11–15 were removed (see Table 1). The remaining data set was missing 6 % of item responses. Little’s Missing Completely at Random (MCAR) test was not significant, $\chi^2(531) = 544.58$, $p = .332$, therefore data were MCAR.

For all parametric analyses throughout the manuscript, variables were checked for skewness. If skewed, variables were checked for outliers, defined as at least 1.5 times the interquartile range above Quartile 3 or below Quartile 1 (Tukey, 1977). Outliers were Winsorized, a common method used to limit the effect of outliers (Field, 2018). In particular, outliers were replaced with values which were 1.5 times the interquartile range above Quartile 3 or below Quartile 1. If variables were still skewed, variables were transformed to best normalize data, following Box and Cox (Box & Cox, 1964; Osborne, 2010), which tended to involve e.g., root, ln, and inverse transformations. Where variables were negatively skewed, variables were reflected before the transformation by subtracting each value from the maximum value plus one, and then reflected back after the transformation by subtracting each value from the maximum of the transformed variable. In Study 1, items 16 and 17 (see Table 1) were positively skewed. All responses to item 17 that were not scored as 2, “Disagree”, were more than 1.5 times the interquartile range above Quartile 3 or below Quartile 1, therefore we did not replace these 92 outliers. These were instead normalized with a square root transformation. Items 3 and 4 were negatively skewed, but had no outliers. Item 3 was normalized with a reflected square root transformation, and item 4 with a reflected 1.5 root transformation. This may reflect that some items were easier, or more encouraged by parents (16, 17), while others were more difficult, or less encouraged by parents (2, 4). Items were correlated. No items were collinear, all *Pearson’s* $r < .61$. Next we examined whether each item (raw or transformed) correlated with the average of all (raw) items using *Pearson’s* $r > .3$, $p < .05$ (Pedhazur & Schmelkin, 1991). The mean EIPSS scores were negatively skewed. Five outliers were replaced, but the scores remained skewed. The outlier-adjusted scores were then normalized with a reflected square root transformation. Item 16 did not correlate with the EIPSS strongly enough, $N = 244$, $r = .15$, so was dropped. Item 17 did correlate just strongly enough, $N = 223$, $r = .33$. However, given the other items correlated much more strongly (above .5, see below), and given the low variability in responses (see above), we chose to delete this item. The average of the remaining items was negatively skewed. Four outliers were replaced, but the scores remained skewed. The outlier-adjusted scores were then normalized with a reflected 1.5 root transformation. All remaining items correlated with the transformed EIPSS scores above the $r = .3$ level, all $N = 234$–265, *Pearson’s* $r > .54$, $p < .001$ (see Table 1 for all correlations). The results remained very similar when outliers were not replaced.

Exploratory Factor Analysis (EFA) was used using Maximum Likelihood, replacing missing values with the mean. The scree plot for Sample 1 shows two factors loaded at eigenvalues above 1 (see Appendix B). Parallel Analysis suggested both factors were above chance levels (see Appendix B). We first ran an EFA with two factors with Varimax rotation, however the average within factor correlation (r = .45) was similar to the between factor correlation (r = .50). Therefore, we next ran the EFA with two factors, using Direct Oblimin rotation, allowing factors to correlate (Samuels, 2017). Based on the Pattern Matrix, which indicates the factor loadings, 5 of the 10 items loaded onto Factor 1 at loadings of .42 or higher, with all 5 items loading best onto this Factor. Factor 1 accounted for 34 % of the variance. The other 5 items loaded onto Factor 2 at loadings of .44 or higher, with all 5 items loading best onto this Factor. Factor 2 accounted for 9 % of the variance. Table 1 shows the factor loadings for each item. Based on the items which scored most strongly onto each Factor, Factor 1 may represent Repetitive Problem Solving (RPS), e.g., building with blocks, sorting objects, or doing puzzles. These are general IPS activities that children might repeat, although they might solve different problems each time, e.g., building a boat, versus a castle, with blocks; or completing a jigsaw in a different order. Factor 2 may represent Novel Problem Solving (NPS), e.g., determining how new toys work, or determining how to fix toys. These IPS activities are generally done only once.

Children's overall RPS scores were averaged across the 5 final items: $M = 2.47$, $Range = 1.00–4.00$, $SD = 0.69$. Children's overall NPS scores were averaged across the 5 final items: $M = 2.72$, $Range = 1.00–4.00$, $SD = 0.63$. Internal reliability analysis using McDonald's omega was good for Factors 1, $\omega = .77$, and 2, $\omega = .73$, using multiple imputation (Hayes & Coutts, 2020). Factors 1 and 2 strongly correlated, *Pearson* $r = 0.50$, $p < .001$. As the correlation between the two subscales was above .30, all items should be considered as one overall EIPSS scale (Clark & Watson, 1995). Yet we should also keep both subscales as the average correlation within subscales ($r = 0.45$) was higher than the average correlation between subscales ($r = 0.29$) (Clark & Watson, 1995). Children's overall EIPSS scores were averaged across the 10 final items: $M = 2.60$, $Range = 1.00–4.00$, $SD = 0.57$. See Appendix C for EIPSS and subscale descriptive statistics by age in years.

2.3. Discussion

Study 1 found 10 of the 17 EIPSS items correlated with the average score and showed good internal reliability. An EFA suggested a 2-factor structure. We propose the first factor links to repetitive IPS skills, where children repeat the IPS process with the same objects, such as completing puzzles or sorting objects. We propose the second factor links to novel IPS skills, where children independently solve a problem only once, e.g., determining how new toys work, or fixing broken toys. Interestingly, neither of the social items loaded onto the EIPSS (items 12 and 16, Table 1) suggesting IPS with objects and people may be different skills. Alternatively, since items 13 and 15 were not about objects either, but instead about making plans, or having questions, perhaps the EIPSS captures object-specific IPS rather than general IPS. Study 2 examined whether the factor structure replicated in a separate sample.

3. Study 2: confirmatory factor analysis

3.1. Method

3.1.1. Participants

See Appendix A for the power analysis. There were 567 children in Study 2. Data were collected from June 2016 to September 2019. While 200 children were required for replication, at least 450 participants were recruited so all three studies would add up to at least 782 for demographics analyses in Study 5 (see Appendix A). Participants were recruited as in Study 1. All participants completed a demographics survey (see Table 2). Rewards for participation were as in Study 1.

3.1.2. Measure

3.1.2.1. EIPSS (final version). Based on Study 1, the second version had 10 items (see Appendix D for the final EIPSS).

3.2. Results

The final data set was missing 8 % of item responses. Little's MCAR test was not significant, $\chi^2(419) = 420.34$, $p = .472$. Items 3, 4, 5, 6, and 7 were negatively skewed. Item 3 was normalized by replacing 35 outliers, and a reflected square root transformation. Item 4 was normalized by replacing 6 outliers, and a reflected 1.25 root transformation. Items 5, 6, and 7 were normalized with reflected 1.25 root transformations.

A CFA was performed to confirm the 2-factor structure found in Study 1 using AMOS 26. Full Information Maximum Likelihood was used to account for missing data (Arbuckle, 2018). Correlations between error terms were allowed for items within factors which

Table 3
Descriptive Statistics for the EIPSS, RPS, NPS, GATTeB, UBT, and Fine and Gross Motor Age.

	EIPSS	RPS	NPS	GATTeB	UBT	Gross Motor Age	Fine Motor Age
N	85	85	85	85	85	85	83
Mean	2.91	2.98	2.85	9.67	26.46	33.44	35.78
Range	2.00–3.60	2.00–3.80	1.80–3.80	1–16	7–41	15–55	19–58
SD	0.34	0.40	0.46	3.81	6.13	9.70	10.46

measured similar skills to improve model fit: items 2 and 9 (determining what to do with toys), items 4 and 7 (understanding the mechanisms underlying toys), and items 5, 6, and 8 (sorting objects). The final CFA model did not show adequate fit in terms of p-value, $N = 567$, $\chi^2(29) = 75.09$, $p < .001$, however, it showed good fit in terms of χ^2/df ratio = 2.59 (< 3), root mean square error of approximation (RMSEA) = .05, comparative fit index (CFI) = .95, and incremental fit index (IFI) = .96. Given the large sample size, this suggests these other markers are adequate to accept the model's fit (Bentler, 1990; Hu & Bentler, 1999). See Table 1 for the standardized regression weights. Additionally, the RPS and NPS correlated moderately to strongly, $r = .48$. Internal reliability was good for the RPS, $\omega = .88$, and the NPS, $\omega = .79$ (Brown, 2015). The results remained very similar when outliers were not replaced.

The mean score for the EIPSS was 2.78 (*range* = 1.33–4.00; *SD* = 0.44; *N* = 567). The mean score for the RPS was 2.65 (*range* = 1.00–4.00; *SD* = 0.64; *N* = 567). The mean score for the NPS was 2.90 (*range* = 1.40–4.00; *SD* = 0.46; *N* = 564). See Table 3 for EIPSS, RPS, and NPS descriptive statistics by age, in years.

3.3. Discussion

A CFA confirmed the 2-factor structure of the EIPSS, with both factors having good internal reliability. Study 3 sought convergent validity between the EIPSS and researcher-led IPS and divergent thinking lab tasks.

4. Study 3: convergent validity

4.1. Method

4.1.1. Participants

See Appendix A for the power analysis. Eighty-five children were tested between March 2017 and July 2018. Participants were recruited through Bounty packs within Sheffield, UK, press releases, and Facebook advertising within Sheffield, UK (see Table 2 for demographics). All children participated, and received a book.

4.2. Measures

4.2.1. EIPSS. Same as study 2

4.2.1.1. Great Ape Tool Test Battery (GATTeB). Reindl et al. (2016) developed the GATTeB to measure IPS with tools in 2- to 3.5-year--olds (in comparison to ape behavior). These tasks were chosen to compare the EIPSS to a direct IPS behavioral task. We used four of the original tasks: Algae Scoop, Seed Extraction, Termite Fish and Marrow Pick (see Fig. 1). These tasks were chosen from the original 12-item battery because they showed a variety of success rates in the original data: 60 % success in Termite Fish, 40 % in Algae Scoop, 29.4 % in Seed Extraction and 17.9 % in Marrow Pick.

In the Algae Scoop task (Fig. 1A), children were asked to remove a black strip of plastic from a box. In the Seed Extraction task

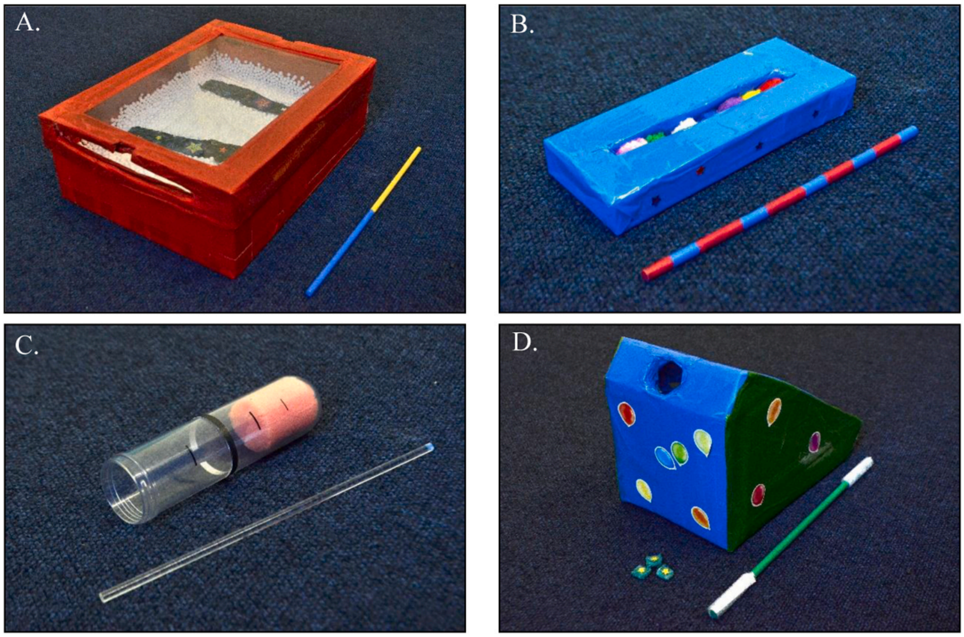


Fig. 1. The 4 tasks chosen from the Great Ape Tool Test Battery (GATTeB). A: Algae Scoop. B: Seed Extraction. C: Marrow Pick. D: Termite Fish.

(Fig. 1B), children were asked to remove a ball. In the Marrow Pick task (Fig. 1C) children were asked to remove the sponge from a tube. In the Termite Fish task (Fig. 1D), children were asked to remove a star.

4.2.1.2. Unusual Box Test (UBT). The UBT (Fig. 2) measures divergent thinking in 1- to 4-year-olds (Bijvoet-van den Berg & Hoicka (2014); Hoicka et al., 2016). This measure was chosen as divergent thinking has been theorized, and shown, to be an important creative process for problem solving (Carr et al., 2016; Guilford, 1975; Kim, 2006; Meadow et al., 1959; Parnes & Meadow, 1959). It is a colorful box with attachments, incorporating strings, blocks, a round hole, rings, stairs, and a rectangular room. Five items are used with the box: a plastic hook, a shaker, a rubber toy, an egg cup, and a spatula. The box was placed on a turntable so that children could easily access the different sides of the box.

4.2.1.3. Child development inventory (CDI) (Ireton, 1992). Since the GATTeB and UBT are both physical tasks, fine and gross motor age were measured as control variables, as these should be more accurate predictors of task performance than chronological age. Parents answered questions about fine and gross motor development with items such as, "Stand alone, steady." They continued answering questions until they answered, "No" three times in a row. Following the CDI coding scheme, children's fine and gross motor age were defined by the age associated with the last item for which parents answered "Yes."

4.2.2. Design

The EIPSS, RPS, NPS, GATTeB, and UBT scores were the main variables in this correlational design. Gross and Fine Motor Age were

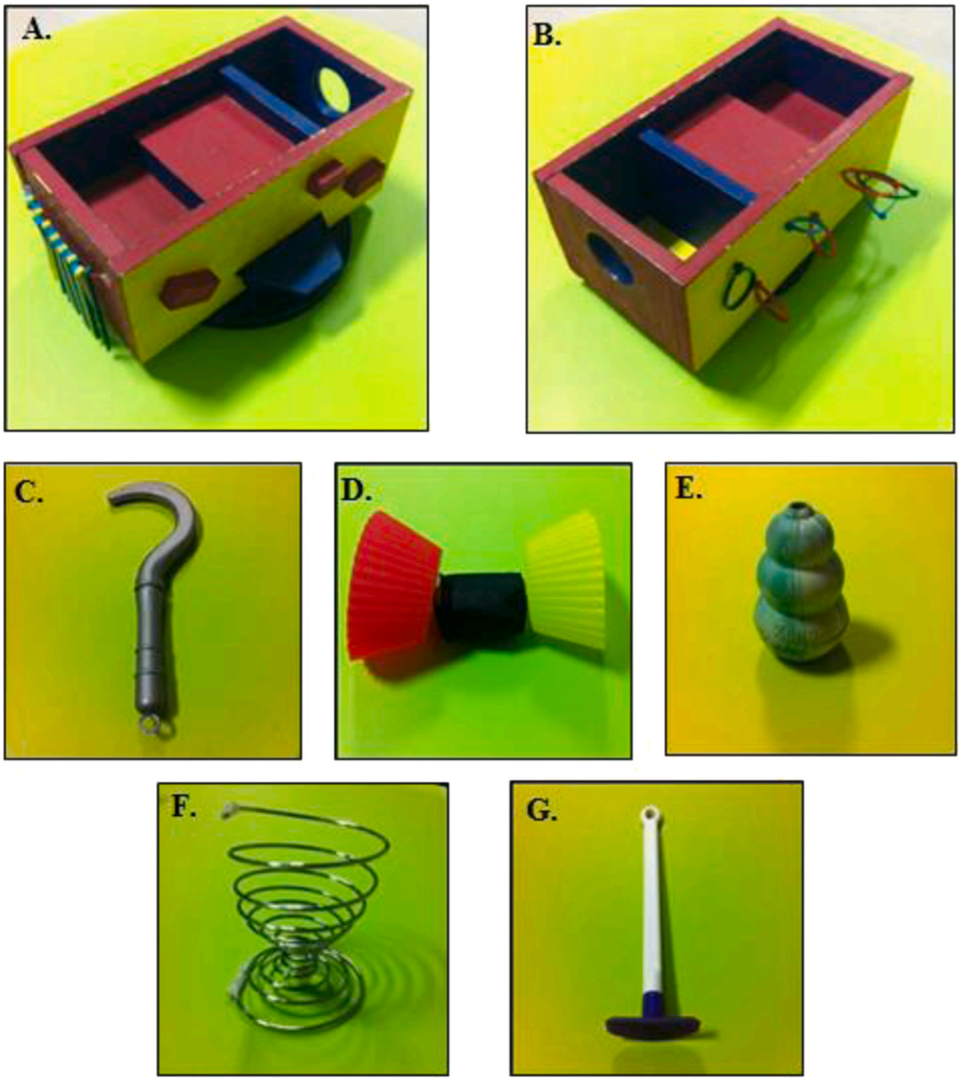


Fig. 2. The Unusual Box Test. A: View of the blocks, rings, and stairs. B: View of the round hole, rings, and rectangular room. C: Plastic hook. D: Shaker. E: Rubber toy. F: Egg cup. G: Spatula.

control variables. The GATTeB and UBT were counterbalanced, and task orders were counterbalanced within each measure. See Bijvoet-van den Berg & Hoicka (2014) for UBT counterbalancing. See Appendix E for GATTeB counterbalancing.

4.2.3. Procedure

Parents completed the EIPSS and CDI. For the GATTeB, the experimenter placed one task in front of the child and asked them to remove the item in the box or tube according to the script (see Appendix E). For example, for the Termite Fish, the experimenter said, “If you can get the stars out the box, you win a sticker!” The child had one minute (Termite Fish and Marrow Pick) or two minutes (Algae Scoop and Seed Extraction) to remove the item from the box or tube. The experimenter never referred to the stick. Therefore, children had to determine for themselves that the stick could help them remove the item. Once the child completed the task, or if they ran out of time, the experimenter gave them the sticker, took away the task, and brought out the proceeding task. Children always received a sticker so they would continue with the tasks instead of becoming upset.

For the UBT, the experimenter showed the different box features using a script (see Bijvoet-van den Berg & Hoicka (2014)). The experimenter then encouraged children to turn the box around. Then the experimenter handed one of the five items to the child. Children had 90 s of freeplay with the box and item. Every 90 s, the experiment gave the child a new item, and the process repeated until all five items were used.

4.2.4. Coding

Tasks were recorded with camcorders filming children from two opposite angles. We coded from video.

4.2.4.1. GATTeB. Children received one point for picking up the stick; and an additional point for using the stick in a way potentially leading to achieving the goal, e.g., in the Marrow Pick task, inserting the stick into the tube and touching the sponge. They received either two points for successfully completing the task with the stick; one point for an alternative method (e.g., using fingers instead of the stick in the Seed Extraction task); or no points if they did not successfully complete the task. Children therefore scored 0–4 points per task, for a total of 0–16 points for all four tasks.

The coding scheme for the GATTeB broadly followed Reindl et al. (2016). However, in the original study, children who completed the task without the tool were coded as “incorrect success” and excluded from analyses. An additional coding feature was added, where children who successfully completed the task without using a tool or according to the desired sequence were coded as “alternative success.” This was to distinguish between children who were not successful at completing the task at all and children who completed the task using an alternative, non-tool-based method. Children scored higher for solving the task with a tool, compared to without a tool, because tool use is a sophisticated behavior revealing greater planning abilities and the ability to understand causality compared to non tool-use behavior (Chen et al., 2010; Keen, 2011; Seed & Call, 2014). For example, in the Algae Scoop, children could put their hand through the hole to reach the plastic, even though their hand cannot comfortably fit in the hole. A better solution is using the stick, saving potential effort and discomfort.

To examine whether children were more likely to pick up, use, and successfully use a tool in later trials (order effects), we used Friedman K analyses. A trend suggested children may have been more likely to pick up a tool in later trials, T1 = 74.1 %; T2 = 71.4 %; T3 = 80.0 %; T4 = 84.7 %; $\chi^2(3) = 6.98$, $p = .073$. Children were significantly more likely to correctly use a tool in later trials, T1 = 63.5 %; T2 = 60.0 %; T3 = 75.3 %; T4 = 82.4 %; $\chi^2(3) = 21.00$, $p < .001$. A trend suggested children were more likely to achieve success with a tool in later trials, T1 = 27.1 %; T2 = 41.2 %; T3 = 34.1 %; T4 = 43.5 %; $\chi^2(3) = 7.13$, $p = .068$. In contrast, children were significantly less likely to achieve success *without* a tool in later trials, T1 = 30.6 %; T2 = 18.8 %; T3 = 12.9 %; T4 = 18.8 %; $\chi^2(3) = 10.40$, $p = .015$. This suggests children may have found tool-based solutions, rather than hand-based solutions, to be more effective over time, supporting our coding scheme which awards more points to successes with tools.

4.2.4.2. UBT. There were 10 box areas (e.g., stairs, hole), and 18 possible actions (e.g., jump, roll), which were combined to code each unique action-box area combination, e.g., using the spatula to “jump on the edge of the box”, “drop in the rectangular room” or “guide through the strings”. All unique action-box area combinations were summed to produce a divergent thinking score. More detailed information about the UBT coding scheme is in Bijvoet-van den Berg & Hoicka (2014).

4.2.4.3. Reliability. To assess reliability for both the GATTeB and UBT, 20 % of the videos ($N = 17$) were coded for agreement by a second coder. Reliability was excellent for the GATTeB (*Intraclass correlation, ICC,* using 2-way mixed effects =.99, $p < .001$) and good for the UBT (*ICC* =.85, $p < .001$).

4.3. Results

See Table 3 for means, ranges, and standard errors for the EIPSS, RPS, NPS, GATTeB, UBT, Fine Motor Age, and Gross Motor Age. One percent of items were missing from the EIPSS. Little’s MCAR test was not significant, $\chi^2(89) = 93.69$, $p = .346$. Item 3 was positively skewed, but was normalized by replacing 2 outliers. The same CFA was run as in Study 2. The final model showed adequate fit, $N = 85$, $\chi^2(29) = 30.13$, $p = .408$, and χ^2/df ratio = 1.04, RMSEA = .02, CFI = .98, and IFI = .99. See Table 1 for standardized regression weights. The RPS and NPS strongly correlated, $r = .52$. Internal reliability was slightly low for the RPS, $\omega = .69$, and good for the NPS, $\omega = .81$ (Brown, 2015). The results remained very similar when outliers were not replaced.

Gross Motor Age was positively skewed, but normalized with a cube root transformation. The EIPSS, RPS, and GATTeB correlated with Gross Motor Age, all $N = 85$, *Pearson's* $r > .26$, $p < .016$. The RPS and UBT correlated with Fine Motor Age, both $N = 83$, *Pearson's* $r > .38$, $p < .001$. Raw correlations were run between the EIPSS and subscales with the GATTeB and UBT. Partial correlations were also run, correcting for Fine and/or Gross Motor Age if they correlated with either of the variables (see earlier this paragraph). Table 4 shows neither the EIPSS, RPS, nor the NPS correlated with either the GATTeB or UBT. However, the NPS had a small to moderate correlation with the UBT when Fine Motor Age was controlled for.

4.4. Discussion

Study 3 showed the NPS correlated with the UBT when fine motor age was controlled for. This suggests parent reports of children's NPS relate to researcher observations of children's divergent thinking in a lab setting, and hence, convergent validity. However, the EIPSS and the RPS did not correlate with the GATTeB or UBT, suggesting limited convergent validity.

It is surprising the UBT correlated with the NPS while the GATTeB did not, since the GATTeB tests IPS more directly than the UBT. Possible reasons for this are discussed in the general discussion. However, this finding backs the idea that divergent thinking, that is, generating many ideas, supports IPS (Carr et al., 2016; Gönül et al., 2019; Guilford, 1975; Kim, 2006; Meadow et al., 1959; Parnes & Meadow, 1959). However, our findings diverge from previous results finding no link between IPS and divergent thinking. While Gönül et al. (2019) found a link between divergent thinking and IPS in 5- to 6-year-old Turkish children, they did not find this link in 5- to 6-year-old children from New Zealand. This link was also not found in British 5- to 7-year-old (Beck et al., 2016), or Congolese 4- to 12-year-olds fdic(Lew-Levy et al., 2021). One difference between our study and these other studies is that our divergent thinking task was physical, while their tasks were verbal (asking about e.g., different ways to use a pipe cleaner). It is therefore possible that a physical divergent thinking task might better predict physical IPS. However, we did not find a correlation between the UBT and the GATTeB, which is also physical. Therefore, it may also be that our survey measure, focusing on more naturalistic forms of novel IPS (e.g., fixing a toy in real life), links to divergent thinking, while lab tasks do not necessarily.

Study 4 sought to demonstrate inter-observer reliability on the EIPSS from both parents, and determine whether parents reported consistent EIPSS scores over a 6-month interval.

5. Study 4: inter-observer reliability and longitudinal stability

5.1. Method

5.1.1. Participants

See Appendix A for power analyses. Parents from Studies 1 and 2 who completed the survey online were automatically invited to have the other parent complete the survey. Data from a second parent were collected from March 2016 to September 2019. Reliability between parents was run for a subsample of participants from Studies 1 and 2 ($N = 32$ children; 18 male, 14 female; *Mean* = 31 months, 18 days; *SD* = 8 months, 7 days; *Range* = 14 months, 4 days – 45 months, 10 days). A subset of parents from Study 2 were invited to repeat the survey 6 months later by email. Time 2 data were collected from October 2018 to March 2019. A test for 6-month longitudinal stability was run for a subsample from Study 2 ($N = 110$; 44 male, 66 female; Time 1 *Mean* = 27 months, 20 days; *SD* = 10 months, 10 days; *Range* = 12 months, 1 day – 47 months, 18 days). Up to £ 2 was donated to charity (e.g., UNICEF), or a £ 5 Amazon voucher (or equivalent in other countries) was sent to parents, for each repeated survey, or for which a second parent completed the survey.

5.2. Measure

5.2.1. EIPSS

Same as Study 2.

Table 4
Pearson Correlations, and Partial Correlations (in Brackets) Controlling for Gross Motor Age (EIPSS, RPS, GATTeB), and Fine Motor Age (RPS, UBT). N = 85, Unless Controlling for Fine Motor Age, for which N = 83.

	GATTeB	UBT
EIPSS	–.01	.18
	(–.09)	(.15)
RPS	.07	.12
	(–.16)	(–.01)
NPS	-.09	.15
	(.03)	(.22)*

*p < .05

5.3. Results

5.3.1. Inter-observer reliability

On average, when both parents ($N = 32$ children) completed the EIPSS, they completed them 4.16 days apart ($SD = 5.63$ days; *Range* = 0–19 days). A trend suggested the EIPSS between parents (*ICC using 1-way Random Effects* =.42, $p = .067$) did not correlate at an acceptable level. There was a good correlation on the RPS, $ICC = .72$, $p < .001$. There was very poor agreement on the NPS, $ICC = –.18$, $p = .675$.

5.3.2. Longitudinal stability

A subsample of parents from Study 2 ($N = 110$) completed the EIPSS on average 6 months and 6 days after previously completing it ($SD = 5$ days; *Range* = 5 months, 27 days – 6 months, 29 days). Outliers on the EIPSS (1), RPS (3), and NPS (4) at time 2 were replaced. The EIPSS scores at times 1 and 2 (*ICC using 2-way random effects* =.83, $p < .001$), the RPS ($ICC = .83$, $p < .001$), and the NPS ($ICC = .69$, $p < .001$) showed good longitudinal stability. Results remained very similar when outliers were not replaced.

5.4. Discussion

Study 4 found the EIPSS and both its subscales had good longitudinal stability after 6 months. This suggests children's propensity for IPS is relatively stable over time. While inter-observer reliability between parents on the RPS was good, it was not very strong for the EIPSS as a whole, and was very poor for the NPS. This suggests that while different parents are consistent in reporting children's IPS involving repetitive tasks, e.g., building structures with blocks, they are not consistent in reporting novel IPS behaviors, e.g., fixing toys. One reason may be that the former may, by its repetitive nature, be more frequently observed than the latter, due to its non-repetitive nature. Thus, repetitive IPS might be more likely to be observed by both parents, while novel IPS might be less likely to be observed by both parents. Study 5 sought to determine whether the EIPSS was useable across different demographic groups, and to implement the EIPPS as a research tool to examine demographic differences.

6. Study 5: demographic similarities and differences

6.1. Method

6.1.1. Participants

See Appendix A for power analyses. All participants from Studies 1–3 were included ($N = 924$).

6.1.2. Measure

6.1.2.1. EIPSS. Same as Study 2. Demographic information was also used, including child: age, gender, mono-/multilingualism, siblings, country; and parent: age, gender, education, ethnicity, and household income (see Table 2). Parents were asked, "Which language(s) is your child regularly exposed to?" Monolingualism was coded if parents reported children were exposed to only one language. Multilingualism was coded if parents reported children were exposed to more than one language.

6.2. Results

6.2.1. Reliability across different demographic groups

Data were pooled from all three samples ($N = 924$). Items 2, 3, 4, 5, 6, and 7 were negatively skewed. Item 4 was normalized by replacing 26 outliers; item 3 by replacing 73 outliers, and using a reflected 1.5 root transformation; items 5 and 6 with a reflected 1.25 root transformation; and items 2 and 7 with a reflected 1.1 root transformation. The average EIPSS scores were negatively skewed, and normalized with a reflected 1.5 root transformation. Age in days was positively skewed, and normalized with a square root transformation. Income (UK) was positively skewed, and normalized by replacing 13 outliers and using a square root transformation. Income (US) was positively skewed and normalized by replacing 1 outlier.

Differential Item Functioning (DIF) was used to determine whether item responses loaded onto the EIPSS differently by key demographic variables. This statistic is commonly used to detect response bias to items (Penfield & Camilli, 2006), including psychological tests using Likert scales (Zumbo, 1999). We used linear regression, with each EIPSS item as the dependent variable, the average EIPSS score as the independent variable in Step 1, and both the demographic variable, and the interaction of the demographic variable and average EIPSS score in Step 2. If there was a significant difference between the models in Steps 1 and 2, the difference in variance explained by each model (the Zumbo-Thomas effect size) was examined. A Zumbo-Thomas effect size above .13 indicates people in different demographic groups respond differently to the item (Zumbo, 1999). Appendix F shows no significant item differences across any demographic variables with Zumbo-Thomas scores above .13, including Child Age in days ($N = 924$), Child Gender ($N = 464$ female; $N = 458$ male), Parent Education (with degree, $N = 713$; without degree, $N = 150$), Income (UK, $N = 426$; US, $N = 84$), Country (UK, $N = 730$, vs. USA, $N = 97$; UK vs. Australia, $N = 25$), Parent Ethnicity (Black, Asian and minority ethnic (BAME) ethnicities, $N = 66$; White ethnicity, $N = 721$), or Parent Gender ($N = 742$ female; $N = 45$ male). This suggests no meaningful item differences for any demographic variables measured. This implies EIPSS items were answered in similar ways across demographic groups. Results did not change when outliers were not replaced.

6.2.2. Changes with child age

Data were pooled across all three samples ($N = 924$) to determine which factors correlated with the EIPSS and its subscales' scores with a small to large effect size, depending on the sample sizes (see Appendix A). This was to examine whether the EIPSS, RPS, and NPS scores changed with age in more detail. Stepwise linear regressions were run on each of the EIPSS (raw score), RPS (raw score), and NPS (raw score) as each of the dependent variables, and Child Age (months, raw), Child Age squared, and Child Age cubed as the independent variables, to examine linear, quadratic, and cubic age changes on the raw EIPSS data. Raw data were used to give a good visual illustration of predicted scored by age (see Appendices G and H). The model for the EIPSS, $N = 924$, $F(2, 921) = 58.27$, $p < .001$, $R^2 = .11$ (moderate effect size), found Child Age, $\beta = .052$, $t = 5.57$, $p < .001$, and Child Age squared, $\beta = –0.01$, $t = –4.04$, $p < .001$, predicted the EIPSS, while Child Age cubed did not. Appendix G suggests EIPSS scores increased with age until around 36 months, then leveled off. The model for the RPS, $N = 920$, $F(2, 918) = 190.89$, $p < .001$, $R^2 = .20$ (large effect size), found Child Age, $\beta = .124$, $t = 10.98$, $p < .001$, and Child Age squared, $\beta = –0.02$, $t = –8.28$, $p < .001$, predicted the RPS, while Child Age cubed did not. Appendix H suggests RPS scores increased with age until around 38 months, and then leveled off. No age variables predicted the NPS.

6.2.3. Demographic differences

Variables were transformed as necessary as described in the previous section on DIF. The following variables were negatively skewed, but normalized by reflected 1.25 root transformations after replacing outliers: RPS (14 outliers), and NPS (15). The following variables were positively skewed, but normalized by 1.25 root transformations after replacing outliers: Childcare Hours (2 outliers), and Parent Age (14). Table 5 shows initial *Pearson* and *Spearman's Rho* correlations between the EIPSS, RPS, NPS, and all demographic variables. *Spearman's Rho* was used when the demographic variable was binary, and in these cases, raw data were used for the parametric variables. When both variables were continuous, transformed variables were used where necessary for the *Pearson* correlations. The EIPSS showed initial positive correlations with both subscales, Child Age (days), and having Siblings. The RPS showed initial positive correlations with the NPS, Child Age, Childcare Hours, and having Siblings, and girls scored significantly higher than boys. The NPS showed initial negative correlations with Parent Education and Childcare, and boys scored significantly higher than girls. The Pearson correlations were similar when outliers were not replaced, except that the correlation between the NPS and Childcare Hours was not significant; there was a trend instead of a significant relationship between Childcare Hours and Parent Age; and there were trends between the EIPSS and Income (US), and Child Age and Income (US), instead of no significant relationships.

Table 5

Initial Correlations (Partial Correlations in Brackets) between the EIPSS, RPS, NPS, and Demographic Variables. Spearman's Rho and Partial Non-Parametric Correlations were Used for Binary Variables, Including Child Gender, Parent Education (with Degree, without Degree), Siblings, Multilingualism, and Parent Gender. Any Correlations not Involving these Variables Used Pearson Correlations and Partial Pearson Correlations. Child Age was Controlled for in Partial Correlations Involving the EIPSS, RPS, Childcare Hours, Siblings, and Parent Age. Child Gender was Controlled for in Partial Correlations Involving the RPS and NPS.

	2	3	4	5	6	7	8	9	10	11	12	13
1. **EIPSS**	.85 *	.77 *	.33 *	–.02	–.06^t	–.02	.17	.02	.08 *	.01	.05	.03
	(.84 *)	(.82 *)	*NA*	(–.04)	(–.08 *)	(–.06)	(.12)	(–.07)	(–.05)	(.01)	(.04)	(–.09 *)
N	921	924	924	922	863	426	84	564	786	783	787	784
2. **RPS**		.33 *	.51 *	.09 *	–.01	–.01	.19^t	.10 *	.13 *	–.01	.06	.04
		(.40 *)	*NA*	(.07 *)	(–.04)	(–.05)	(.12)	(–.01)	(–.05)	(–.03)	(.04)	(–.09 *)
N		921	921	919	860	425	83	561	783	780	785	781
3. **NPS**			–.01	–.15 *	–.08 *	–.04	.06	–.07 *	–.02	.04	.01	.00
			NA	*NA*	(–.08 *)	(–.05)	(.09)	(–.08 *)	(–.04)	(.03)	(.01)	(–.04)
N			924	922	863	426	84	564	786	783	787	784
4. **Child Age (Days)**				.04	.02	.07	.16	.19 *	.28 *	.02	.04	.17 *
N				922	863	426	84	564	786	783	787	784
5. **Child Gender**					.02	–.03	.16	–.04	–.01	–.04	.02	.00
N					861	424	84	562	784	781	785	782
6. **Parent Education**						.35 *	.41 *	.25 *	–.01	.03	.05	.19 *
N						421	84	551	769	765	776	773
7. **Income (UK)**								.35 *	-.04	–.10 *	.08	.35 *
N								348	424	420	424	423
8. **Income (USA)**								.50 *	–.04	.20^t	.33 *	.41 *
N								83	84	83	84	83
9. **Childcare Hours**									–.09 *	.07	.00	.09 *
N									560	560	554	554
10. **Siblings**										–.09	.05	.18 *
N										776	775	772
11. **Multilingualism**											.04	.04
N											772	769
12. **Parent Gender**												.02
N												780
13. **Parent Age**												

Girls scored as 1, boys scored as – 1; *$p < .05$, $^t p < .10$

Table 5 shows partial correlations for the EIPSS, RPS, and NPS with all other variables. The control variables were Child Age and Child Gender for any variables showing initial correlations with these variables (see Table 5, and previous paragraph). Partial correlations were non-parametric when the demographic variable was binary, and in these cases, the raw data were used for the continuous variables. When both main variables were continuous, transformed variables were used where necessary. The EIPSS showed positive partial correlations with its subscales, and negative partial correlations with Parent Education and Parent Age, when Child Age was controlled for. The RPS showed girls scored significantly higher than boys when Child Age was controlled for. There was also a positive partial correlation with the NPS, and a negative partial correlation with Parent Age, when Child Age and Child Gender were controlled for. The NPS showed a negative partial correlation with Parent Education when Child Gender was controlled for, and a negative partial correlation with Childcare Hours when Child Gender and Child Age were controlled for. The partial Pearson correlations were similar when outliers were not replaced, except there were trends between the EIPSS and Income (UK), and the NPS and Parent Age, instead of no significant relationships.

To determine which demographic variables best predicted the EIPSS, a stepwise linear regression was run on the EIPSS (transformed) as the dependent variable, Child Age (transformed) as an independent variable in Step 1 (as it correlated with the EIPSS, see Table 5), and Parent Education and Parent Age (transformed) as independent variables in Step 2 (as these partially correlated with the EIPSS, see Table 5). The model was improved by Child Age, $F(1, 771) = 199.81$, $p < .001$, $\Delta R^2 = .21$, followed by Parent Education, $F(1, 770) = 6.40$, $p = .012$, $\Delta R^2 = .01$, but not Parent Age. The final model found Child Age, $t = 14.22$, $p < .001$, $\beta = .02$, and Parent Education, $t = -2.53$, $p = .012$, $\beta = -.03$, were both significant. Therefore, more educated parents, but not older parents, reported lower EIPSS scores, once Child Age was controlled for. When outliers were not replaced, we found Parent Age predicted the EIPSS in addition to Child Age and Parent Education. However, we accept the initial result as it is more conservative.

To determine which demographic variables best predicted the NPS, a stepwise linear regression was run on the NPS (transformed) as the dependent variable, Child Gender as an independent variable in Step 1 (as it correlated with the NPS, see Table 5), and Parent Education and Childcare Hours (transformed) as independent variables in Step 2 (as these partially correlated with the NPS, see Table 5). The model was improved by Child Gender, $F(1, 547) = 15.37$, $p < .001$, $\Delta R^2 = .03$, followed by Parent Education, $F(1, 546) = 5.39$, $p = .021$, $\Delta R^2 = .01$, but not Childcare Hours. The final model found Child Gender, $t = -3.80$, $p < .001$, $\beta = -.12$, and Parent Education, $t = -2.32$, $p = .021$, $\beta = -.05$, were both significant. Therefore, more educated parents, but not parents of children who had more childcare hours, reported lower NPS scores, once Child Gender was controlled for. Results were very similar when outliers were not replaced. As the RPS only correlated with one demographic variable (Parent Age) once Child Gender was controlled for (as Child Gender correlated with the RPS), a regression analysis was not run.

ANCOVA was used to examine the effects of Country (Australia, Canada, UK, USA) on the EIPSS (transformed), with Child Age (transformed) as a covariate (as it correlated with the EIPSS, see Table 5). Regression slopes between the EIPSS (transformed) and Child Age (transformed) were all similar within each country (Australia *Pearson* $r = .54$; Canada $r = .38$; UK $r = .32$; USA $r = .42$). An initial ANCOVA determined there was no interaction between Country and Child Age (transformed), $F(3, 860) = 1.83$, $p = .139$, therefore this interaction was dropped from the final model. The final model did not violate Levene's test of equality of variances, $F(3, 864) = 1.65$, $p = .177$. The model was significant, $F(4, 863) = 31.00$, $p < .001$, $\eta_p^2 = .13$. Child Age, $F(1, 863) = 112.01$, $p < .001$, $\eta_p^2 = .12$, and Country, $F(3, 863) = 3.88$, $p = .009$, $\eta_p^2 = .01$, both improved the model. However, posthoc pairwise comparisons with Sidak corrections found no significant differences between countries, all $p > .152$.

ANCOVA was used to examine the effects of country (Australia, Canada, UK, USA) on the RPS (transformed), with Child Age (transformed) and Child Gender as covariates (as they correlated with the RPS, see Table 5). Regression slopes between the RPS (transformed) and Child Age (transformed) were all similar within each country (Australia *Pearson* $r = .54$; Canada $r = .65$; UK $r = .50$; USA $r = .58$), as were those between the RPS (transformed) and Child Gender (Australia: *Spearman's rho* $r = .34$; Canada $r = .38$; UK $r = .07$; USA $r = .18$). An initial ANCOVA determined there was no interaction between Country and Child Age (transformed), $F(3, 851) = 1.37$, $p = .250$, nor between Country and Child Gender, $F(3, 851) = 0.26$, $p = .857$, therefore these interactions were dropped from the final model. The final model did not violate Levene's test of equality of variances, $F(3, 859) = 1.01$, $p = .389$. The model was significant, $F(5, 857) = 64.06$, $p < .001$, $\eta_p^2 = .27$. Child Age, $F(1, 857) = 305.09$, $p < .001$, $\eta_p^2 = .26$, and Child Gender, $F(1, 857) = 5.10$, $p = .024$, $\eta_p^2 = .01$, were both significant. However, Country, $F(3, 857) = 1.25$, $p = .292$, was not. Results were very similar when outliers were not replaced.

ANCOVA was used to examine the effects of country (Australia, Canada, UK, USA) on the NPS (transformed), with Child Gender as a covariate (as it correlated with the NPS, see Table 5). Regression slopes between the NPS (transformed) and Child Gender were different across countries, with Australia and Canada showing positive slopes, and the UK and USA showing negative slopes (Australia: *Spearman's rho* $r = .17$; Canada $r = .31$; UK $r = -.15$; USA $r = -.23$). However, an initial ANCOVA determined there was no interaction between Country and Child Gender, $F(3, 858) = 2.09$, $p = .100$, therefore the interaction was dropped from the final model. The final model did not violate Levene's test of equality of variances, $F(3, 862) = 0.01$, $p = .999$. The model was significant, $F(4, 861) = 6.62$, $p < .001$, $\eta_p^2 = .03$. Child Gender, $F(1, 861) = 13.03$, $p < .001$, $\eta_p^2 = .02$, and Country, $F(3, 861) = 4.03$, $p = .007$, $\eta_p^2 = .01$, were both significant. Posthoc pairwise comparisons with Sidak corrections found no significant differences between countries, all $p > .067$. Results were very similar when outliers were not replaced.

6.3. Discussion

No meaningful item functioning differences existed across the EIPSS in relation to child age, child gender, parent education, household income (UK or USA), country (UK vs. USA; UK vs. Australia), parent ethnicity, or parent gender. Therefore, the EIPSS may be useful across these demographic groups. However, only child age, child gender, and household income (UK) were powered for a small effect size. Therefore, caution should be taken in accepting the results for the other demographic variables, as meaningful significant differences may exist with larger samples. In future, DIF should be examined across these, and other (e.g., Canada) demographic groups when powering for a small effect size.

When looking at predicted age curves, the EIPSS and RPS both increased with Child Age (months), but decreased with Child Age squared. However, the NPS did not change with age. Several other demographic differences were found. First, girls had significantly higher RPS scores, while boys had significantly higher NPS scores (once child age was controlled for the RPS only, both small effect sizes). Second, children scored higher on the EIPSS and the NPS if their parent *did not* have a degree (once child age was controlled for the EIPSS, and child gender was controlled for the NPS, small effect sizes). Finally, children scored higher on the NPS when parents were younger (once child gender was controlled for, small effect size).

7. General discussion

These studies found the 10-item EIPSS may be a useful measure of IPS for children between 12 and 47 months. The EIPSS showed a 2-factor structure (Repetitive, Novel) across three separate samples, where both factors correlated. The EIPSS also showed good longitudinal stability across a 6-month interval. However, inter-observer reliability between parents was not very good, nor was convergent validity with two lab tasks: the GATTeB and UBT. When digging deeper into the subscales, the RPS also showed good longitudinal stability at 6 months' time, and additionally, good inter-observer reliability between parents, although it did not correlate with either lab task. The NPS also showed good longitudinal stability, but poor inter-observer reliability between parents. However, it correlated with a lab task – the UBT, suggesting convergent validity. The results altogether suggest some aspects of the EIPSS and its subscales are valid and reliable, while others are not. Therefore, caution should be taken in using the EIPSS and its subscales in future research. Further research may need to be done to ensure full validity and reliability, and subscales may need to be formally dissociated (see 7.1. Limitations and Future Directions).

Furthermore, the EIPSS showed no item functioning differences across child age, parent education (with or without a degree), income (within the UK or USA), country (UK, USA, Australia), parent ethnicity, or parent gender, suggesting it works across a variety of demographic settings. However, caution should be taken as several of these measures did not power for small effect sizes (e.g., country, parent ethnicity), thus small item functioning differences may in theory exist. Future research should examine this possibility with larger sample sizes.

As children aged, they scored higher on the EIPSS until around 36 months, and the RPS drove these results, which increased with age until around 38 months. This suggests children may develop repetitive IPS skills (e.g., sorting, building, solving puzzles) as their cognitive or motor abilities advance. Alternatively, perhaps these activities are regularly available at home, and so can be practiced repeatedly, with children performing more complex activities (e.g., creating more complex builds with blocks) over time. However, the fact that the RPS no longer increases after 38 months is surprising given past research found children's Repetitive IPS increases with age after 3 years (Montford & Readdick, 2008). Future research should consider how to better tap into age, e.g., adding more difficult RPS items. In contrast, the NPS did not correlate with age. This is surprising as Novel IPS increases with age in general (Babik et al., 2019; Barrett et al., 2007; Breyel & Pauen, 2021; Gardiner et al., 2012; Gönül et al., 2018; Hopper et al., 2020; Neldner et al., 2019; Rat-Fischer et al., 2014; Rat-Fischer et al., 2012; Reindl et al., 2016; Seed & Call, 2014; Tarasuik et al., 2017; Willatts, 1984, 1999). This may, therefore, suggest a validity issue with the NPS, as it does not track age as we might expect. One reason we might see age differences for the RPS, but not the NPS, is that RPS items generally captured success (e.g., Item 1, "My child makes difficult structures with blocks/megablocks on his/her own"), while NPS items generally captured drive (e.g., Item 7, "My child looks at the mechanisms of how things work, e.g., how wheels are connected to toy cars.") This difference might be an alternative explanation for why the EIPSS has two subscales, which could potentially be considered "success" and "drive" rather than "novel" and "repetitive". By wording questions in these ways, the RPS might be able to capture age changes in success, while the NPS might instead capture children's general attitude to RPS, which might be more of an individual difference, like e.g., temperament (Putnam et al., 2006). In future, it would be interesting to examine whether the changing of the wording might lead the RPS items to be less age-dependent (e.g., "My child *likes to* make difficult structures with blocks/megablocks on his/her own"), and to make NPS items more age-dependent (e.g., "My child *understands* the mechanisms of how things work, e.g., how wheels are connected to toy cars.").

Several other demographic factors correlated with the EIPSS and its subscales. EIPSS and NPS scores were higher for children whose parents *did not* have a degree, in contrast to null and opposite results in past research (Day & Burns, 2011; Greiff et al., 2013; Hacatrjana, 2022; Levine et al., 2012). While no child gender differences existed across the EIPSS as a whole, girls scored higher on the RPS, diverging from past research finding girls and boys did equally well at solving puzzles, and that girls were *more* likely to ask for help than boys (Thompson, 1999; Thompson & Moore, 2000). In contrast, boys scored higher on the NPS, converging with results finding boys were more likely than girls to use a tool to reach out-of-reach objects by themselves (Gredlein & Bjorklund, 2005), and that older (5- to 6-year-old boys) are more likely to innovate tools than girls in Turkey, although not in New Zealand (Gönül et al., 2019). Children with older parents scored lower on the NPS.

One possible explanation for some of these results is rooted in scaffolding. Research suggests parents scaffold their children's problem solving abilities, building children's problem solving skills until they can solve problems independently (Mermelshtine, 2017). However, counterintuitively, children receiving less scaffolding in day to day life may be more used to discovering solutions for themselves, which in many cases is a less effective solution for task success (Rendell et al., 2010), but in the case of *independent* problem solving may provide an advantage as such children may practice innovation more, and may therefore need less help to solve these tasks. For instance, parents from higher SES backgrounds, including higher levels of education and household income, increase their scaffolding with young children (Carr & Pike, 2012; Dilworth-Bart et al., 2010; Lowe et al., 2014). However, parents *without* degrees reported their children had higher EIPSS scores, and in particular, higher Novel subscale scores. In the case of parent age, older parents scaffold more than younger parents (Wood et al., 2016), yet in the current paper, older parents reported *lower* RPS scores. Similarly, in the case of gender, parents scaffolded for girls more in the first half of a block tower-building task, and more for boys in the second half of the task (Conner & Cross, 2003). Girls are also more likely to ask to watch others demonstrate solving a novel problem compared to boys (Rawlings et al., 2021) which may give girls more scaffolding opportunities than boys. Thus, boys might be better at NPS items because they are used to initially figuring out how to do novel tasks on their own. Yet, this does not explain why girls would score higher on the RPS. However, this is speculative, and future research should examine the involvement of parental scaffolding in detail to determine whether it explains, at least in part, these demographic differences. Other explanations should also be researched, such as the availability of different types of toys and objects to children in these different demographic groups; the possibility of biological differences in the case of gender (Gredlein & Bjorklund, 2005); or the possibility parents observe and report differently about boys and girls, or report differently depending on their own SES or age. For example, parents with higher levels of education might have higher expectations for their children's IPS, and so may not recognize, and therefore not report, simpler forms of IPS on the EIPSS. For instance, on item 1, “My child makes difficult structures with blocks/megablocks on his/her own”, parents with higher levels of education may give lower scores for a simple tower, while parents with lower levels of education may give higher scores for the same simple tower.

7.1. Limitations and future directions

The EIPSS has room for improvement. While no DIF differences existed across countries (UK, USA, Australia) these analyses did not test for small effect sizes, and the Canadian sample was not large enough to compare Canada to the UK. Future research should not only compare large enough samples from different countries to examine small effect sizes, but should also interview parents across different countries to ensure the items make sense in different cultural contexts (DeVellis, 2017). For instance, item 1, “My child makes difficult structures with blocks/megablocks on his/her own” uses the word “blocks.” While American and Canadian parents use the word “blocks”, British parents tend to favor the word “bricks” instead, possibly causing confusion for British parents.

A second limitation is that the EIPSS showed inter-observer reliability on the RPS only. One reason inter-observer reliability may have worked in this way is that, by its nature, a repeated event, compared to a novel event, is more easily observed on a day-to-day basis. If one of the parents in each pair is a primary caregiver, and the second is not, spending much less time with their child, they may have had less chance to observe the novel behaviors, leading to poor inter-observer reliability. To obtain inter-observer reliability on the NPS, one may need to compare scores of people who spend large amounts of time with the same children, e.g., a primary caregiver working part-time, and a key worker at a childcare setting.

A third limitation is that we found convergent validity on the NPS only. Mumford et al. (1991) suggest creative problem solving taps into several cognitive processes which build up in stages. At the beginning of the process, people use *problem construction*, where they try to figure out what the problem is in the first place; and *information encoding*, in which they seek to find information out about the problem. These processes may be key for both the UBT and the NPS. The UBT is an open problem, in which children must come up with problems to solve through encoding information about the box. Similarly, items on the NPS focus on children discovering problems, and encoding information about them, e.g., item 7, “My child looks at the mechanisms of how things work, e.g., how wheels are connected to toy cars”. In contrast, the RPS may tap into mid-level creative problem solving processes of *categorization*, in which children figure out how to match the problem to past knowledge, for instance, item 5, “My child sorts objects in logical ways on his/her own, e.g., cutlery next to plates; toy bed in dollhouse”. Finally, the GATTeB is focused on the later *implementation* stage of problem solving, where children receive points for attempting to complete, and successfully completing, the task. To obtain convergent validity, it may work to test children on a series of lab tasks that directly parallel the items in the EIPSS. For instance, Hoicka et al. (2021) matched specific socio-cognitive lab tasks to specific survey items (e.g., gaze-following) on the Early Social Cognition Inventory and found convergent validity. Future research could compare children's EIPSS scores to lab tasks in which children perform tasks which match the items on the EIPSS, e.g., puzzles, construction, fixing a toy, or getting an out-of-reach toy. Such a lab task may better match the underlying cognitive processes being used in the EIPSS (Mumford et al., 1991), and may also be more representative of the types of IPS children perform in real life. A further limitation of the GATTeB, in terms of measuring IPS, is there is, to some extent, a social element to the task. The experimenter highlights the goal, e.g., getting the stars out of the box. Thus, children do not have to identify the problem. Future tasks should therefore leave children alone with the objects without instruction to see how they do at, not only solving the problem, but also identifying the problem, as they may do in real life. Given the limitations of the EIPSS in converging with the lab task, at this point in time, it is not prudent to use the EIPSS as a control measure for IPS lab tasks. Furthermore, given the differences in validity and reliability for the NPS and RPS, it may also be best to use the subscales only, rather than the EIPSS as a whole.

Finally, it is worth noting that many of the individual items were negatively skewed across each sample. Interestingly, it appears that the younger the mean age of the sample, the more individual items were negatively skewed. For instance, the youngest sample, in Study 2 (mean age 26 months, 24 days) had 5 negatively skewed items. The eldest sample, in Study 3 (mean age 36 months, 13 days) had only 1 negatively skewed item. Finally, Study 1 (mean age 28 months; 5 days) had 2 negatively skewed items. This may then reflect that the younger children were, the less likely they were to demonstrate IPS for a behavior. Furthermore, item 3, “When it is physically difficult to reach a toy, my child will not try to figure out how to reach it” was negatively skewed across all three samples, suggesting it may be a relatively rare type of IPS compared to the other types. However, it can be beneficial to have some items at the extremes of a scale to widen the response rates (Clark & Watson, 1995).

With further refinement, the EIPSS may be valuable for future research on IPS in 1- to 3-year-olds. By running the EIPSS, which takes approximately 3 min to complete, alongside other surveys and experiments, researchers could efficiently learn how IPS relates to a variety of skills in early development. Given the example of comparing social learning and innovation discussed in the introduction (Bonawitz et al., 2011; Carr et al., 2015; Hoicka et al., 2018; Legare & Nielsen, 2015; Rawlings et al., 2017; Rawlings et al., 2021), the EIPSS could be run alongside a social learning lab task to easily learn more about this relationship in the early years. Using the EIPSS alongside other surveys could be an easy, economical way to start to understand how IPS relates to e.g., social cognition (Hoicka et al., 2021; Tahiroglu et al., 2014), cognition (Baker et al., 2013), language (Fenson et al., 1994), temperament (Putnam et al., 2006), and parenting styles (Winstanley & Gattis, 2013), among other constructs. Survey results could then lead onto experiments and interventions to investigate those constructs which are best linked to IPS. The EIPSS could also be used longitudinally to examine how early IPS links to later skills. For instance, Pásztor et al. (2015) found creative thought processes, in this case divergent thinking, predicted mathematical achievement in 6th-graders. Glover (1979) found more creative university students (based on divergent thinking) asked more abstract questions (involving evaluation and synthesis) in relation to texts they read, compared to less creative students, who focused their questions on facts. Similarly, one could examine the EIPSS's relationship to later mathematical achievement or reading comprehension in primary school.

The EIPSS could also be useful in a variety of practical settings. Early years educators, medical staff, and parents could use the EIPSS as a guideline to what constitutes appropriate types of early IPS from 1 to 3 years. For instance, some items from the EIPSS could be included in early years curricula which focus on problem solving (Australian Government Department of Education and Training, 2017; Best Start Expert Panel on Early Learning, 2007; Department for Education, 2017; Ohio Department of Education, 2012), e.g., building structures and puzzles, sorting objects, and understanding toys' underlying mechanisms. Given that adolescents with Down syndrome and Williams syndrome have difficulties with IPS (Camp et al., 2016), it may be beneficial to use the EIPSS with young children with these developmental differences to see how they develop compared to children with typical development.

7.2. Conclusions

The EIPSS is the first parent-report measure of IPS in children aged 12–47 months. It shows a good factor structure, revealing dimensions of both repetitive and novel IPS. The EIPSS provides a potential tool to better understand how early IPS develops (e.g., through cognitive or motor development), and how it affects other areas of development (e.g., IPS in domains such as mathematics or language) across English-speaking countries. However more research is needed to determine whether the EIPSS is valid and reliable in terms of inter-observer reliability, and convergent reliability.

Declaration of Interest

Elena Hoicka owns the survey website www.babylovesscience.com that collected the data. There are no other conflicts of interest.

Data availability

The authors do not have permission to share data.

Acknowledgments

We thank parents and children for participating.

Funding: This work was supported by a Returning Carers' Scheme grant from the University of Bristol awarded to the first author; an ESRC PhD studentship awarded to the second author; and an Applied Psychology Centre temporary fund grant, Staffordshire University, for the project 'The Immediate Impact of Different Types of Television on Young Children's Creativity' awarded to the third author.

Open Practices Statement: The datasets generated during the current study are available from the authors on reasonable request, in collaboration with the authors.

Appendix A

Power analyses for all studies.

Analysis	Statistic	N^{req}	N^{act}	Source
Study 1: Survey Construction	EFA	170	272	17 items; 10 participants per item (Tabachnick & Fidell, 2007)
Study 2: Survey Replication	CFA	200	567	Minimum 200 participants (Kline, 2011)
Study 3: Concurrent Validity	Correlation	84	85	2-tailed medium correlation ($r = 0.3$, based on previous surveys) (Libertus & Landa, 2013; Winstanley & Gattis, 2013), with $\alpha = 0.05$, power = 0.8 (Faul et al., 2007)

(*continued on next page*)

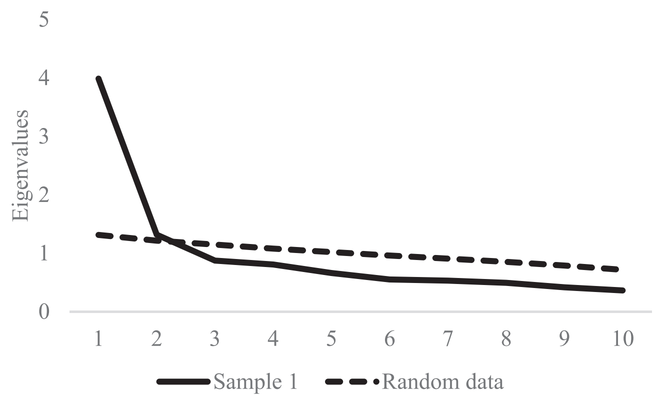
(*continued*)

Analysis	Statistic	N^{req}	N^{act}	Source
Study 4: Inter-observer Reliability	Intraclass Correlation	10	32	2-tailed very large correlation (*ICC* = 0.7) with 2 observations per participant $\alpha = 0.05$, power = 0.8 (Bujang & Baharum, 2017)
Study 4: Longitudinal Stability	Correlation	29	110	2-tailed large correlation (*r* = 0.5, based on previous surveys) (Putnam et al., 2006), with $\alpha = 0.05$, power = 0.8 (Faul et al., 2007)
Study 5: Child Age; Parent Income (UK)	Differential Item Functioning	400	≥ 426	Small effect size (based on simulations, corrections for multiple testing) (Belzak, 2020)
Study 5: Child Gender	Differential Item Functioning	200/ group	≥ 458/ group	Small effect size (based on simulations, corrections for multiple testing) (Belzak, 2020)
Study 5: Parent Education (Degree, No Degree)	Differential Item Functioning	100/ group	≥ 150/ group	Medium effect size (based on simulations, no corrections for multiple testing) (Belzak, 2020)
Study 5: Income (US)	Differential Item Functioning	50	84	Large effect size (based on simulations, no corrections for multiple testing) (Belzak, 2020)
Study 5: Country (UK, USA); Country (UK, Australia); Parent Ethnicity (BAME, White)	Differential Item Functioning	25/ group	≥ 25/ group	Large effect size (based on simulations, no corrections for multiple testing) (Belzak, 2020)
Study 5: Child Age, Parent Age	Pearson Correlation	782	≥ 781	Small effect size (r = .10) with $\alpha = 0.05$, power = 0.8 (Faul et al., 2007)
Study 5: Child Gender, Parent Education, Siblings, Multilingualism, Parent Gender	Spearman's Rho Correlation	779	≥ 780	Small effect size (*r* = .10) with $\alpha = 0.05$, power = 0.8 (Faul et al., 2007)
Study 5: Childcare Hours, Income (UK)	Pearson Correlation	193	≥ 425	Small to medium effect size (*r* = .20) with $\alpha = 0.05$, power = 0.8 (Faul et al., 2007)
Study 5: Income (US)	Pearson Correlation	46	≥ 83	Medium to large effect size (*r* = .40) with $\alpha = 0.05$, power = 0.8 (Faul et al., 2007)
Study 5: Country (Australia, Canada, UK, US)	ANCOVA	13/ group	≥ 15/ group	Large effect size (Cohen's *f* =.40) with $\alpha = 0.05$, power = 0.8 (Faul et al., 2007)

N^{req} is the minimum number of participants required. N^{act} is the actual number of participants in the sample for each analysis. For Study 5, analyses were posthoc for certain variables (e.g., US income), as these groups were not targeted *a priori*.

Appendix B

Parallel Analysis for Study 1: Eigenvalues for Sample 1 and random data.



.

Appendix C

Descriptive statistics for age (years), EIPSS, RPS, and NPS, by year, in Studies 1–3; as well as the GATTeB and UBT in Study 3.

Age:	1	2	3
Study 1*N*	119	73	80
Age *M*	18;15	29;5	41;17
Range	12;4-23;28	24;1-35;28	36;3-47;26
SD	3;10	3;19	3;24
EIPSS *M*	2.59	2.56	2.66

(*continued on next page*)

(*continued*)

Age:	1	2	3
Range	1.56-3.67	1.00-3.90	1.00-4.00
SD	0.39	0.62	0.73
RPS *M*	2.30	2.54	2.67
Range	1.00-3.60	1.00-4.00	1.00-4.00
SD	0.54	0.64	0.84
NPS *M*	2.87	2.58	2.64
Range	1.80-3.80	1.00-3.80	1.00-4.00
SD	0.43	0.73	0.73
Study 2*N*	251*	192*	124
Age *M*	17;16	29;12	41;15
Range	12;0-23;27	24;0-35;23	36;2-47;23
SD	3;10	3;12	3;13
EIPSS *M*	2.56	2.93	3.01
Range	1.33-3.60	1.89-4.00	1.80-3.80
SD	0.42	0.38	0.35
RPS *M*	2.23	2.92	3.09
Range	1.00-3.80	1.40-4.00	1.60-4.00
SD	0.59	0.46	0.45
NPS *M*	2.86	2.93	2.93
Range	1.80-4.00	1.40-4.00	1.80-4.00
SD	0.47	0.46	0.44
Study 3*N*	*NA*	43	42
Age *M*		30;8	42;23
Range		24;20-35;28	36;9-47;25
SD		3;16	3;10
EIPSS *M*		2.84	3.00
Range		2.00-3.60	2.40-3.60
SD		0.34	0.31
RPS *M*		2.85	3.11
Range		2.00-3.80	2.40-3.60
SD		0.42	0.34
NPS *M*		2.83	2.87
Range		1.80-3.80	2.00-3.60
SD		0.50	0.41
GATTeB *M*		8.72	10.64
Range		1-16	2-16
SD		3.36	4.04
UBT *M*		24.21	28.76
Range		12-34	7-41
SD		5.24	6.17

*N = 249 for the RPS for 1-year-olds in Study 2; N = 191 for the RPS for 2-year-olds in Study 2.

Appendix D

Final EIPSS.

Instructions: For each statement, think back to specific examples in the last month, and choose whether you: Strongly Agree; Agree; Disagree; Strongly Disagree. However, sometimes you will not be able to think of a specific time in the last month to evaluate the statement. For instance, if we ask about how your child responded to getting a present, and your child didn't get a present in the last month, skip the question or choose NA (not applicable).

Item	Question
1	My child makes difficult structures with blocks/megablocks on his/her own
2 [R]	My child needs help from others when figuring out how new toys work
3 [R]	When it is physically difficult to reach a toy, my child will not try to figure out how to reach it
4 [R]	My child does not like to take things apart and put them back together
5	My child sorts objects in logical ways on his/her own, e.g., cutlery next to plates; toy bed in dollhouse
6	My child puts away objects correctly on his/her own
7	My child looks at the mechanisms of how things work, e.g., how wheels are connected to toy cars
8 [R]	My child needs help to sort objects by, e.g., shape, colour
9	If a toy breaks, my child tries to fix it before asking for help
10 [R]	My child needs help completing simple puzzles

[R]Item is reverse-scored

Appendix E

Great Ape Tool Test Battery counterbalancing and script.
Counterbalancing:

Order	Task 1	Task 2	Task 3	Task 4
1	Algae Scoop	Seed Extraction	Marrow Pick	Termite Fish
2	Seed Extraction	Algae Scoop	Termite Fish	Marrow Pick
3	Marrow Pick	Termite Fish	Seed Extraction	Algae Scoop
4	Termite Fish	Marrow Pick	Seed Extraction	Algae Scoop
5	Algae Scoop	Seed Extraction	Terappmite Fish	Marrow Pick
6	Seed Extraction	Algae Scoop	Marrow Pick	Termite Fish
7	Marrow Pick	Termite Fish	Algae Scoop	Seed Extraction
8	Termite Fish	Marrow Pick	Algae Scoop	Seed Extraction

Script:.

Category	Script
Introducing GATTeB	"We've got some games to play."
Algae Scoop	"Here is our first/ next game. We have a box and inside the box are some stickers. If you can get the stickers out of the box, you will win a sticker!"
Seed Extraction	"Here is our first/ next game. We have a little box here and inside there are some colorful balls. If you can get the balls out of the box, you'll win a sticker!"
Marrow Pick	"Here is our first/ next game. We have a tube here and inside the tube there's a sponge. If you can get the sponge out of tube box, you'll win a sticker!"
Termite Fish	"Here is our first/ next game. We have a box here and I have these stars. Look what I do – one, two, three! If you can get the stars out the box, you'll win a sticker!"
Successful Task Completion	"Well done, you got the sticker/ star/ sponge out! Now you can choose a sticker!"
Unsuccessful Task Completion	"That one is a bit tricky isn't it? Don't worry, you had a really good try so you can choose a sticker while I get the next game out!"
End of GATTeB	"We've finished playing games now, well done! Have you had fun?"
Additional Prompts	"You're doing really well" "Keep trying" "Can you try?" "Good job!"
If child asks for help	"I'm not sure how to do it, what do you think?"

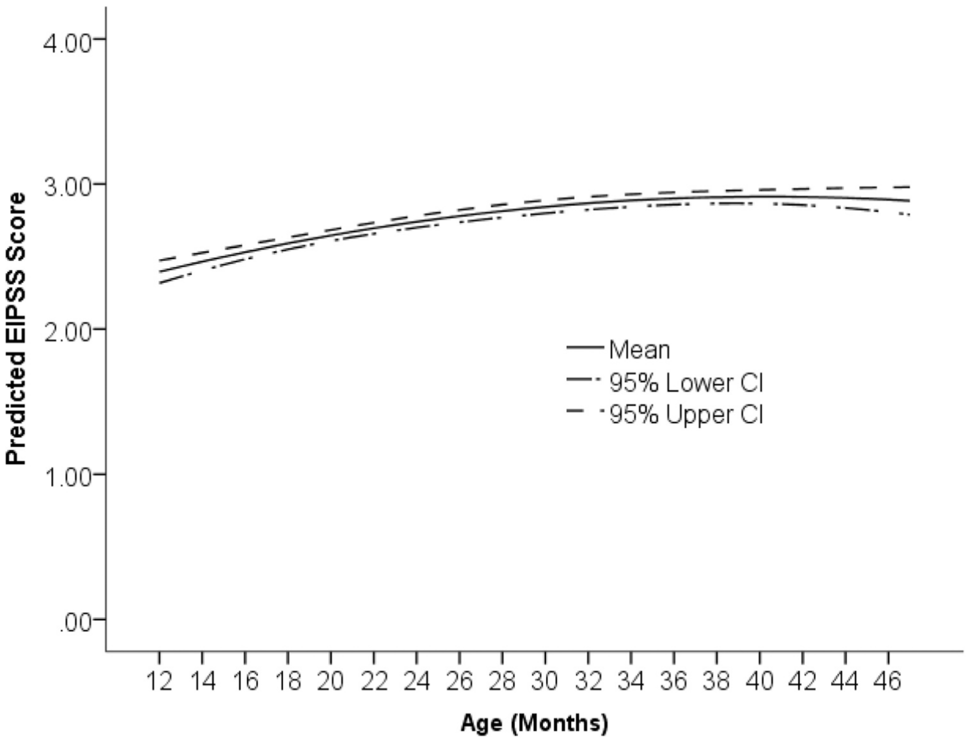
Appendix F

Differential Item Functioning for Child Age in days (*N* = 924), Child Gender (*N* = 464 female; *N* = 458 male), Parent Education (with degree, *N* = 713; without degree, *N* = 150), Income (UK, *N* = 426; USA, *N* = 84), Country (UK, *N* = 730, vs. USA, *N* = 97; UK vs. Australia, *N* = 25), Parent Ethnicity (BAME ethnicities, *N* = 66; White ethnicity, *N* = 721); and Parent Gender (*N* = 742 female; *N* = 45 male). ΔR^2 are Zumbo-Thomas effect sizes. Significant p-values are .005 for Child Age, Child Gender, and Income (UK) to account for Bonferroni corrections. Significant p-values are .05 for Country, Parent Education, Parent Income (USA), Parent Ethnicity (BAME, White), and Parent Gender, to account for smaller sample sizes, as simulations suggest correcting for multiple tests in small samples is not advisable for DIF (Belzak, 2020). *NA* = Not Applicable, as p-values were not significant.

Item	Child Age		Child Gender		Parent Education		Income (UK)		Income (US)		UK vs US		UK vs Australia		Parent Ethnicity		Parent Gender	
	p	ΔR^2	*p*	ΔR^2	*P*	ΔR^2	*p*	ΔR^2	*p*	ΔR^2	*p*	ΔR^2	*p*	ΔR^2	*p*	ΔR^2	*p*	ΔR^2
1	**< .001**	**.04**	.928	*NA*	.478	*NA*	.551	*NA*	.578	*NA*	.935	*NA*	.940	*NA*	.609	*NA*	.465	*NA*
2	**.003**	**.01**	.027	*NA*	.132	*NA*	.037	*NA*	.467	*NA*	**.037**	**.01**	.688	*NA*	**.006**	**.01**	.704	*NA*
3	**< .001**	**.04**	.990	*NA*	.978	*NA*	.116	*NA*	.527	*NA*	.997	*NA*	**.018**	**.01**	.594	*NA*	.392	*NA*
4	**< .001**	**.09**	.270	*NA*	.288	*NA*	.455	*NA*	.952	*NA*	.261	*NA*	.445	*NA*	.606	*NA*	.651	*NA*
5	**0.001**	**.01**	.208	*NA*	.856	*NA*	.644	*NA*	.840	*NA*	.210	*NA*	.368	*NA*	.582	*NA*	.415	*NA*
6	.053	*NA*	.826	*NA*	.828	*NA*	.934	*NA*	.966	*NA*	.833	*NA*	.395	*NA*	.831	*NA*	.506	*NA*
7	**< .001**	**.04**	.247	*NA*	.693	*NA*	.509	*NA*	.120	*NA*	.297	*NA*	.672	*NA*	.787	*NA*	.067	*NA*
8	**< .001**	**.09**	.801	*NA*	.584	*NA*	.599	*NA*	.615	*NA*	.795	*NA*	.330	*NA*	**.004**	**.01**	.761	*NA*
9	**< .001**	**.01**	.107	*NA*	.193	*NA*	.344	*NA*	**.041**	**.07**	.124	*NA*	.170	*NA*	**.003**	**.01**	.827	*NA*
10	**< .001**	**.04**	.308	*NA*	.138	*NA*	.233	*NA*	.661	*NA*	.310	*NA*	.291	*NA*	.278	*NA*	.915	*NA*

Appendix G

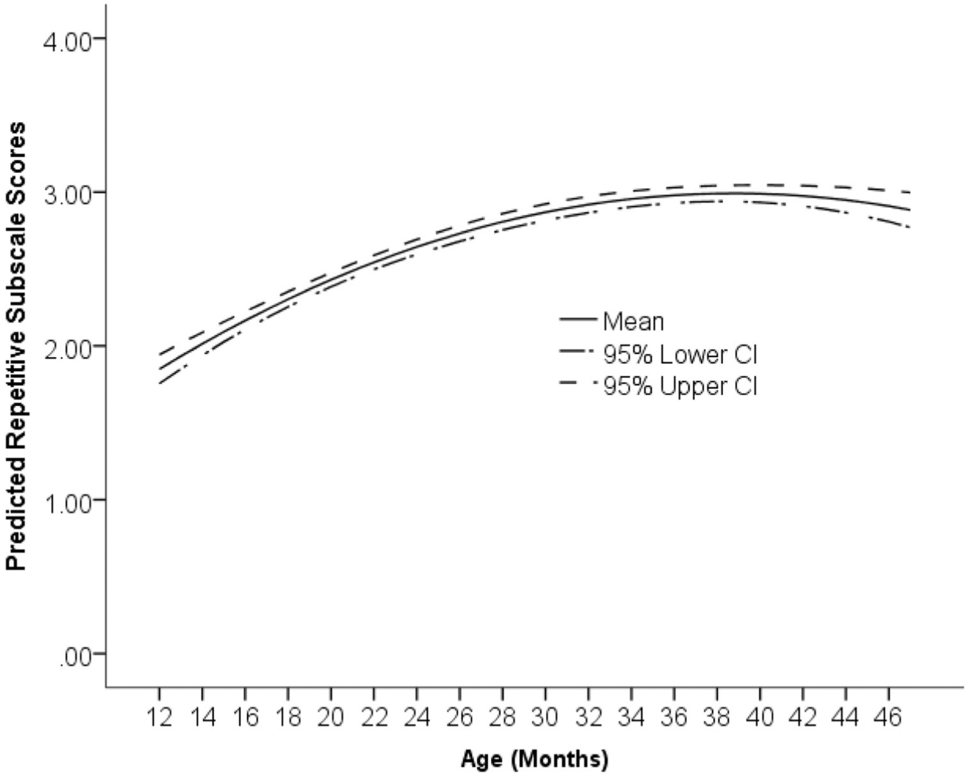
Unstandardized Predicted Values for the EIPSS by Age.



.

Appendix H

Unstandardized Predicted Values for the RPS by Age.



.

References

Adamson, M., & Blight, E. J. (2014). Bringing dads to the table: Comparing mother and father reports of child behaviour and parenting at mealtimes. *Journal of Family Studies, 20*(2), 118–127. https://doi.org/10.1080/13229400.2014.11082001

Alakortes, J., Fyrstén, J., Carter, A. S., Moilanen, I. K., & Ebeling, H. E. (2015). Finnish mothers' and fathers' reports of their boys and girls by using the Brief Infant-Toddler Social and Emotional Assessment (BITSEA). *Infant Behavior and Development, 39*, 136–147. https://doi.org/10.1016/j.infbeh.2015.02.016

Arbuckle, J. L. (2018). *AMOS (version 24.0)*. SPSS Inc.,.

Australian Government Department of Education and Training. (2017). *Belonging, being & becoming: The early years learning framework in Australia*.

Babik, I., Cunha, A. B., Ross, S. M., Logan, S. W., Galloway, J. C., & Lobo, M. A. (2019). Means-end problem solving in infancy: Development, emergence of intentionality, and transfer of knowledge. *Developmental Psychobiology, 61*(2), 191–202. https://doi.org/10.1002/dev.21798

Baker, M., Schafer, G., Alcock, K. J., & Bartlett, S. (2013). A parentally administered cognitive development assessment for children from 10 to 24 months. *Infant Behavior and Development, 36*(2), 279–287. https://doi.org/10.1016/j.infbeh.2013.01.007

Barrett, T. M., Davis, E. F., & Needham, A. (2007). Learning about tools in infancy. *Developmental Psychology, 43*(2), 352–368. https://doi.org/10.1037/0012-1649.43.2.352

Bates, E., Carlson-Luden, V., & Bretherton, I. (1980). Perceptual aspects of tool using in infancy. *Infant Behavior and Development, 3*, 127–140. https://doi.org/10.1016/S0163-6383(80)80017-8

Bechtel, S., Jeschonek, S., & Pauen, S. (2013). How 24-month-olds form and transfer knowledge about tools: The role of perceptual, functional, causal, and feedback information. *Journal of Experimental Child Psychology, 115*(1), 163–179. https://doi.org/10.1016/j.jecp.2012.12.004

Beck, S., Williams, C., Cutting, N., Apperly, I., & Chappell, J. (2016). Individual differences in children's innovative problem-solving are not predicted by divergent thinking or executive functions. *Royal Society of London Proceedings B Biological Sciences, 371*(1690), 20150190. https://doi.org/10.1098/rstb.2015.0190

Belzak, W. C. M. (2020). Testing differential item functioning in small samples. *Multivariate Behavioral Research, 55*(5), 722–747. https://doi.org/10.1080/00273171.2019.1671162

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Best Start Expert Panel on Early Learning. (2007). *Early learning for every child today: A framework for Ontario early childhood settings*.

Bialystok, E., & Majumder, S. (1998). The relationship between bilingualism and the development of cognitive processes in problem solving. *Applied Psycholinguistics, 19*(1), 69–85. https://doi.org/10.1017/S0142716400010584

Bijvoet-van den Berg, S., & Hoicka, E. (2014). Individual differences and age-related changes in divergent thinking in toddlers and preschoolers. *Developmental Psychology, 50*(6), 1629–1639. https://doi.org/10.1037/a0036131

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E., & Schulz, L. (2011). The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition, 120*(3), 322–330. https://doi.org/10.1016/j.cognition.2010.10.001

Bono, M. A., & Stifter, C. A. (2003). Maternal attention-directing strategies and infant focused attention during problem solving. *Infancy, 4*(3), 235–250. https://doi.org/10.1207/S15327078IN0402_05
Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society Series B (Methodological), 26*(2), 211–252. https://doi.org/jstor.org/stable/2984418.
Breyel, S., & Pauen, S. (2021). The beginnings of tool innovation in human ontogeny: How three-to five-year-olds solve the vertical and horizontal tube task. *Cognitive Development, 58*, Article 101049. https://doi.org/10.1016/j.cogdev.2021.101049
Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Publications,.
Bujang, M. A., & Baharum, N. (2017). A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: A review. *Archives of Orofacial Science, 12*(1), 1–11.
Burns, M. S., Haywood, H. C., & Delclos, V. R. (1987). Young children's problem-solving strategies: An observational study. *Journal of Applied Developmental Psychology, 8*(1), 113–121. https://doi.org/10.1016/0193-3973(87)90024-4
Call, J., Carpenter, M., & Tomasello, M. (2005). Copying results and copying actions in the process of social learning: Chimpanzees (Pan troglodytes) and human children (Homo sapiens. *Animal cognition, 8*(3), 151–163. https://doi.org/10.1007/s10071-004-0237-8
Camp, J. S., Karmiloff-Smith, A., Thomas, M. S., & Farran, E. K. (2016). Cross-syndrome comparison of real-world executive functioning and problem solving using a new problem-solving questionnaire. *Research in Developmental Disabilities, 59*, 80–92. https://doi.org/10.1016/j.ridd.2016.07.006
Carr, A., & Pike, A. (2012). Maternal scaffolding behavior: Links with parenting style and maternal education. *Developmental Psychology, 48*(2), 543–551. https://doi.org/10.1037/a0025888
Carr, K., Kendal, R. L., & Flynn, E. G. (2015). Imitate or innovate? Children's innovation is influenced by the efficacy of observed behaviour. *Cognition, 142*, 322–332. https://doi.org/10.1016/j.cognition.2015.05.005
Carr, K., Kendal, R. L., & Flynn, E. G. (2016). Eureka!: What Is Innovation, how does it develop, and who does it. *Child Development, 87*(5), 1505–1519. https://doi.org/10.1111/cdev.12549
Chen, Y., Keen, R., Rosander, K., & Von Hofsten, C. (2010). Movement planning reflects skill level and age changes in toddlers: Movement planning in toddlers. *Child Development, 81*(6), 1846–1858. https://doi.org/10.1111/j.1467-8624.2010.01514.x
Chen, Z., Sanchez, R. P., & Campbell, T. (1997). From beyond to within their grasp: The rudiments of analogical problem solving in 10-and 13-month-olds. *Developmental Psychology, 33*(5), 790–801. https://doi.org/10.1037/0012-1649.33.5.790
Chen, Z., & Siegler, R. S. (2013). Young children's analogical problem solving: Gaining insights from video displays. *Journal of Experimental Child Psychology, 116*(4), 904–913. https://doi.org/10.1016/j.jecp.2013.08.009
Cicirelli, V. G. (1976). Mother-child and sibling-sibling interactions on a problem-solving task. *Child Development, 47*(3), 588–596. https://doi.org/10.2307/1128172
Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309–319.
Conner, D. B., & Cross, D. R. (2003). Longitudinal analysis of the presence, efficacy and stability of maternal scaffolding during informal problem-solving interactions. *British Journal of Developmental Psychology, 21*(3), 315–334. https://doi.org/10.1348/026151003322277720
Cox, D. W. (1985). The Purdue elementary problem-solving inventory (PEPSI), grade level, and socioeconomic status: A preliminary study. *Gifted Child Quarterly, 29*(2), 72–73. https://doi.org/10.1177/001698628502900205
Cushen, P. J., & Wiley, J. (2011). Aha! Voila! Eureka! Bilingualism and insightful problem solving. *Learning and Individual Differences, 21*(4), 458–462. https://doi.org/10.1016/j.lindif.2011.02.007
Davé, S., Nazareth, I., Senior, R., & Sherr, L. (2008). A comparison of father and mother report of child behaviour on the Strengths and Difficulties Questionnaire. *Child Psychiatry and Human Development, 39*, 399–413. https://doi.org/10.1007/s10578-008-0097-6
Day, C. A., & Burns, B. M. (2011). Characterizing the achievement motivation orientation of children from low- and middle-income families. *Early Education and Development, 22*(1), 105–127. https://doi.org/10.1080/10409280903544397
Dean, L. G., Kendal, R. L., Schapiro, S. J., Thierry, B., & Laland, K. N. (2012). Identification of the social and cognitive processes underlying human cumulative culture. *Science, 335*(6072), 1114–1118. https://doi.org/10.1126/science.1213969
Department for Education. (2017). *Statutory Framework for the Early years Foundation stage: Setting the standards for Learning, Development and care for Children from birth to five*.
DeVellis, R. F. (2017). *Scale development: Theory and applications* (*4th ed*..,). Sage,.
Dilworth-Bart, J., Poehlmann, J., Hilgendorf, A. E., Miller, K., & Lambert, H. (2010). Maternal scaffolding and preterm toddlers' visual-spatial processing and emerging working memory. *Journal of Pediatric Psychology, 35*, 209–220. https://doi.org/10.1093/jpepsy/jsp048
Fagard, J., Rat-Fischer, L., & Kevin O'Regan, J. (2014). The emergence of use of a rake-like tool: A longitudinal study in human infants. *Frontiers in Psychology, 5*, 1–12. https://doi.org/10.3389/fpsyg.2014.00491
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146
Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), i-185. https://doi.org/10.2307/1166093
Fidler, D. J., Philofsky, A., Hepburn, S. L., & Rogers, S. J. (2005). Nonverbal requesting and problem-solving by toddlers with Down syndrome. *American Journal on Mental Retardation, 110*(4), 312–322. https://doi.org/10.1352/0895-8017(2005)110[312:NRAPBT]2.0.CO;2
Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed..,). Sage.,.
Flynn, E., Turner, C., & Giraldeau, L. (2016). Selectivity in social and asocial learning: Investigating the prevalence, effect and development of young children's learning preferences. *Philosophical Transactions of the Royal Society B-Biological Sciences, 371*(1690), 20150189. https://doi.org/10.1098/rstb.2015.0189
Gardiner, A. K., Bjorklund, D. F., Greif, M. L., & Gray, S. K. (2012). Choosing and using tools: Prior experience and task difficulty influence preschoolers' tool-use strategies. *Cognitive Development, 27*(3), 240–254. https://doi.org/10.1016/j.cogdev.2012.05.001
Glover, J. A. (1979). Levels of questions asked in interview and reading sessions by creative and relatively noncreative college students. *The Journal of Genetic Psychology, 135*(1), 103–108. https://doi.org/10.1080/00221325.1979.10533421
Gluck, S. C., Tahiroglu, D., & Moses, L. J. (2021). A comparison of mother and father reports of children's theory of mind: Further validation of the children's social understanding scale. *European Journal of Developmental Psychology, 18*(3), 459–469. https://doi.org/10.1080/17405629.2020.1813100
Gönül, G., Hohenberger, A., Corballis, M., & Henderson, A. M. (2019). Joint and individual tool making in preschoolers: From social to cognitive processes. *Social Development, 28*(4), 1037–1053. https://doi.org/10.1111/sode.12373
Gönül, G., Takmaz, E. K., Hohenberger, A., & Corballis, M. (2018). The cognitive ontogeny of tool making in children: The role of inhibition and hierarchical structuring. *Journal of Experimental Child Psychology, 173*, 222–238. https://doi.org/10.1016/j.jecp.2018.03.017
Gredlein, J. M., & Bjorklund, D. F. (2005). Sex differences in young children's use of tools in a problem-solving task. *Human Nature, 16*(2), 211–232. https://doi.org/10.1007/s12110-005-1004-5
Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts—Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology, 105*(2), 364–379.
Guilford, J. P. (1975). Varieties of creative giftedness, their measurement and development. *Gifted Child Quarterly, 19*(2), 107–121. https://doi.org/10.1177/001698627501900216
Hacatrjana, L. (2022). Flexibility to change the solution: An indicator of problem solving that predicted 9th grade students' academic achievement during distance learning, in parallel to reasoning abilities and parental education. *Journal of Intelligence, 10*(1), 7. https://doi.org/10.3390/jintelligence10010007
Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures, 14*(1), 1–24. https://doi.org/10.1080/19312458.2020.1718629
Henderson, B. B., & Dias, L. (1987). An exploratory study of infant problem solving in natural environments. *Ethology & Sociobiology, 8*(3), 205–213. https://doi.org/10.1016/0162-3095(87)90044-6

Hoicka, E., Mowat, R., Kirkwood, J., Carberry, M., Kerr, T., & Bijvoet-van den Berg, S. (2016). One-year-olds think creatively, just like their parents. *Child Development, 87*(4), 1099–1105. https://doi.org/10.1111/cdev.12531

Hoicka, E., Powell, S., Knight, J., & Norwood, M. (2018). Two-year-olds can socially learn to think divergently. *British Journal of Developmental Psychology, 36*(1), 22–36. https://doi.org/10.1111/bjdp.12199

Hoicka, E., Soy Telli, B., Prouten, E., Leckie, G., Browne, W. J., Nurmsoo, E., & Gattis, M. (2021). The Early Social Cognition Inventory (ESCI): An Examination of its Psychometric Properties from Birth to 47 Months. *Behavior Research Methods, 54*, 1200–1226. https://doi.org/10.3758/s13428-021-01628-z

Hopper, L. M., Jacobson, S. L., & Howard, L. H. (2020). Problem solving flexibility across early development. *Journal of Experimental Child Psychology, 200*(104966), 1–18. https://doi.org/10.1016/j.jecp.2020.104966

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Ireton, H. R. (1992). *Child Development Inventory Manual*. Behavior Science Systems,.

Keen, R. (2011). The development of problem solving in young children: A critical cognitive skill. *Annual Review of Psychology, 62*, 1–21. https://doi.org/10.1146/annurev.psych.031809.130730

Kim, K. H. (2006). Is creativity unidimensional or multidimensional? Analyses of the Torrance Tests of Creative Thinking. *Creativity Research Journal, 18*(3), 251–259. https://doi.org/10.1207/s15326934crj1803_2

Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford,.

Krasnor, L. R., & Rubin, K. H. (1983). Preschool social problem solving: Attempts and outcomes in naturalistic interaction. *Child Development, 54*(6), 1545–1558. https://doi.org/10.2307/1129818

Legare, C. H., & Nielsen, M. (2015). Imitation and innovation: The dual engines of cultural learning. *Trends in Cognitive Sciences, 19*(11), 688–699. https://doi.org/10.1016/j.tics.2015.08.005

Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology, 48*(2), 530–542. https://doi.org/10.1037/a0025913

Lew-Levy, S., Pope, S. M., Haun, D. B., Kline, M. A., & Broesch, T. (2021). Out of the empirical box: A mixed-methods study of tool innovation among Congolese BaYaka forager and Bondongo fisher-farmer children. *Journal of Experimental Child Psychology, 211*, Article 105223. https://doi.org/10.1016/j.jecp.2021.105223

Libertus, K., & Landa, R. J. (2013). The Early Motor Questionnaire (EMQ): A parental report measure of early motor development. *Infant Behavior & Development, 36*(4), 833–842. https://doi.org/10.1016/j.infbeh.2013.09.007

Lowe, J., Erickson, S. J., MacLean, P., Duvall, S. W., Ohls, R. K., & Duncan, A. F. (2014). Associations between maternal scaffolding and executive functioning in 3 and 4 year olds born very low birth weight and normal birth weight. *Early Human Development, 90*(10), 587–593. https://doi.org/10.1016/j.earlhumdev.2014.07.009

Meadow, A., Parnes, S. J., & Reese, H. (1959). Influence of brainstorming instruction and problem sequence on a creative problem solving test. *Journal of Applied Psychology, 43*(6), 413–416. https://doi.org/10.1037/h0043917

Mermelshtine, R. (2017). Parent–child learning interactions: A review of the literature on scaffolding. *British Journal of Educational Psychology, 87*(2), 241–254. https://doi.org/10.1111/bjep.12147

Metz, K. E. (1993). Preschoolers' developing knowledge of the pan balance: From new representation to transformed problem solving. *Cognition and Instruction, 11*(1), 31–93. https://doi.org/10.1207/s1532690xci1101_2

Montford, E. I. P., & Readdick, C. A. (2008). Puzzlemaking and part-whole perception of two-year-old and four-year-old children. *Early Child Development and Care, 178*(5), 537–550. https://doi.org/10.1080/03004430600852056

Mumford, M. D., Mobley, M. I., Reiter-Palmon, R., Uhlman, C. E., & Doares, L. M. (1991). Process analytic models of creative capacities. *Creativity Research Journal, 4*(2), 91–122. https://doi.org/10.1080/10400419109534380

Neel, R. S., Jenkins, Z. N., & Meadows, N. (1990). Social problem-solving behaviors and aggression in young children: A descriptive observational study. *Behavioral Disorders, 16*(1), 39–51. https://doi.org/10.1177/019874299001600106

Neldner, K., Mushin, I., & Nielsen, M. (2017). Young children's tool innovation across culture: Affordance visibility matters. *Cognition, 168*, 335–343. https://doi.org/10.1016/j.cognition.2017.07.015

Neldner, K., Redshaw, J., Murphy, S., Tomaselli, K., Davis, J., Dixson, B., & Nielsen, M. (2019). Creation across culture: Children' s tool innovation is influenced by cultural and developmental factors. *Developmental Psychology, 55*(4), 877–889. https://doi.org/10.1037/dev0000672

Neldner, K., Reindl, E., Tennie, C., Grant, J., Tomaselli, K., & Nielsen, M. (2020). A cross-cultural investigation of young children's spontaneous invention of tool use behaviours. *Royal Society Open Science, 7*(5), Article 192240. https://doi.org/10.1098/rsos.192240

Ohio Department of Education. (2012). *Ohio Early Learning and Development Standards Domain: Approaches toward Learning*.

Osborne, J. W. (2010). Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation, 15*(12), 1–9. https://doi.org/10.7275/qbpc-gk17

Parnes, S. J., & Meadow, A. (1959). Effects of "brainstorming" instructions on creative problem solving by trained and untrained subjects. *Journal of Educational Psychology, 50*(4), 171–176. https://doi.org/10.1037/h0047223

Pásztor, A., Molnár, G., & Csapó, B. (2015). Technology-based assessment of creativity in educational context: The case of divergent thinking and its relation to mathematical achievement. *Thinking Skills and Creativity, 18*, 32–42. https://doi.org/10.1016/j.tsc.2015.05.004

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated analysis*. Erlaum,.

Penfield, R. D., & Camilli, G. (2006). Differential Item Functioning and Item Bias. In C. R. Rao, & S. Sinharay (Eds.), *Handbook of statistics* (vol. 26, pp. 125–167). Science Direct. https://doi.org/10.1016/S0169-7161(06)26005-X.

Pepler, D. J., & Ross, H. S. (1981). The effects of play on convergent and divergent problem solving. *Child Development, 52*(4), 1202–1210. https://doi.org/10.2307/1129507

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The early childhood behavior questionnaire. *Infant Behavior and Development, 29*(3), 386–401. https://doi.org/10.1016/j.infbeh.2006.01.004

Rat-Fischer, L., O'Regan, J. K., & Fagard, J. (2012). The emergence of tool use during the second year of life. *Journal of Experimental Child Psychology, 113*(3), 440–446. https://doi.org/10.1016/j.jecp.2012.06.001

Rat-Fischer, L., O'Regan, J. K., & Fagard, J. (2014). Comparison of active and purely visual performance in a multiple-string means-end task in infants. *Cognition, 133*(1), 304–316. https://doi.org/10.1016/j.cognition.2014.06.005

Rat-Fischer, L., O'Regan, J. K., & Fagard, J. (2012). The emergence of tool use during the second year of life. *Journal of Experimental Child Psychology, 113*(3), 440–446. https://doi.org/10.1016/j.jecp.2012.06.001

Rawlings, B., Flynn, E., & Kendal, R. L. (2017). To copy or to innovate? The role of personality and social networks in children's learning strategies. *Child Development Perspectives, 11*(1), 39–44. https://doi.org/10.1111/cdep.12206

Rawlings, B. S., Flynn, E. G., & Kendal, R. L. (2021). Personality predicts innovation and social learning in children: Implications for cultural evolution. *Developmental Science*, e13153). https://doi.org/10.1111/desc.13153

Redding, R. E., Morgan, G. A., & Harmon, R. J. (1988). Mastery motivation in infants and toddlers: Is it greatest when tasks are moderately challenging. *Infant Behavior and Development, 11*(4), 419–430. https://doi.org/10.1016/0163-6383(88)90003-3

Reindl, E., Beck, S. R., Apperly, I. A., & Tennie, C. (2016). Young children spontaneously invent wild great apes' tool-use behaviours. *Proceedings of the Royal Society B: Biological Sciences, 283*(1825), 20152402. https://doi.org/10.1098/rspb.2015.2402

Reindl, E., Gwilliams, A. L., Dean, L. G., Kendal, R. L., & Tennie, C. (2020). Skills and motivations underlying children's cumulative cultural learning: Case not closed. *Palgrave Communications, 6*(1), 1–9. https://doi.org/10.1057/s41599-020-0483-7

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., Fogarty, L., Ghirlanda, S., Lillicrap, T., & Laland, K. N. (2010). Why copy others? Insights from the Social Learning Strategies Tournament. *Science, 328*(5975), 208–213. https://doi.org/10.1126/science.1184719

Rose, S. E., Lamont, A. M., & Reyland, N. (2021). Watching television in a home environment: Effects on children's attention, problem solving and comprehension. *Media Psychology*, 1–26. https://doi.org/10.1080/15213269.2021.1901744

Samuels, P. (2017). *Advice on exploratory Factor Analysis*.

Seed, A. M., & Call, J. (2014). Space or physics? Children use physical reasoning to solve the trap problem from 2.5 years of age. *Developmental Psychology, 50*(7), 1951–1962. https://doi.org/10.1037/a0036695

Sigman, M., Cohen, S. E., Beckwith, L., & Topinka, C. (1987). Task persistence in 2-year-old preterm infants in relation to subsequent attentiveness and intelligence. *Infant Behavior and Development, 10*(3), 295–305. https://doi.org/10.1016/0163-6383(87)90018-X

Speidel, R., Zimmermann, L., Green, L., Brito, N. H., Subiaul, F., & Barr, R. (2021). Optimizing imitation: Examining cognitive factors leading to imitation, overimitation, and goal emulation in preschoolers. *Journal of Experimental Child Psychology, 203*(105036), 1–22. https://doi.org/10.1016/j.jecp.2020.105036

Squires, J., & Bricker, D. (2009). *Ages & Stages Questionnaires®, Third Edition (ASQ®-3): A Parent-Completed Child Monitoring System*. Paul H. Brookes Publishing Co., Inc.,.

Stafford, K. R. (1968). Problem solving as a function of language. *Language and Speech, 11*(2), 104–112. https://doi.org/10.1177/002383096801100203

Subiaul, F., Krajkowski, E., Price, E. E., & Etz, A. (2015). Imitation by combination: Preschool age children evidence summative imitation in a novel problem-solving task. *Frontiers in Psychology, 6*, 1410. https://doi.org/10.3389/fpsyg.2015.01410

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. Allyn & Bacon/Pearson Education,.

Tahiroglu, D., Moses, L. J., Carlson, S. M., Mahy, C. E., Olofson, E. L., & Sabbagh, M. A. (2014). The Children's Social Understanding Scale: Construction and validation of a parent-report measure for assessing individual differences in children's Theories of Mind. *Developmental Psychology, 50*(11), 2485–2497. https://doi.org/10.1037/a0037914

Tarasuik, J., Demaria, A., & Kaufman, J. (2017). Transfer of problem solving skills from touchscreen to 3D model by 3- to 6-year-olds. *Frontiers in Psychology, 8*(1586), 1–6. https://doi.org/10.3389/fpsyg.2017.01586

Tennie, C., Greve, K., Gretscher, H., & Call, J. (2010). Two-year-old children copy more reliably and more often than nonhuman great apes in multiple observational learning tasks. *Primates, 51*(4), 337–351. https://doi.org/10.1007/s10329-010-0208-4

Thompson, R. B. (1999). Gender differences in preschoolers' help-eliciting communication. *The Journal of Genetic Psychology: Research and Theory on Human Development, 160*(3), 357–368. https://doi.org/10.1080/00221329909595405

Thompson, R. B., Cothran, T., & McCall, D. (2012). Gender and age effects interact in preschoolers' help-seeking: Evidence for differential responses to changes in task difficulty. *Journal of Child Language, 39*(5), 1107–1120. https://doi.org/10.1017/S030500091100047X

Thompson, R. B., & Moore, K. (2000). Collaborative speech in dyadic problem solving: Evidence for preschool gender differences in pragmatic development. *Journal of Language and Social Psychology, 19*(2), 248–255. https://doi.org/10.1177/0261927×00019002004

Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley,.

Vlachou, M., & Farrell, P. (2000). Object mastery motivation in pre-school children with and without disabilities. *Educational Psychology, 20*(2), 167–176. https://doi.org/10.1080/713663715

Vygotsky, L. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press,.

Willatts, P. (1984). The Stage-IV infant's solution of problems requiring the use of supports. *Infant Behavior and Development, 7*(2), 125–134. https://doi.org/10.1016/S0163-6383(84)80053-3

Willatts, P. (1999). Development of means–end behavior in young infants: Pulling a support to retrieve a distant object. *Developmental Psychology, 35*(3), 651–667. https://doi.org/10.1037/0012-1649.35.3.651

Winstanley, A., & Gattis, M. (2013). The baby care questionnaire: A measure of parenting principles and practices during infancy. *Infant Behavior and Development, 36*(4), 762–775. https://doi.org/10.1016/j.infbeh.2013.08.004

Wood, E., Petkovski, M., De Pasquale, D., Gottardo, A., Evans, M. A., & Savage, R. S. (2016). Parent scaffolding of young children when engaged with mobile technology. *Frontiers in Psychology, 7*(690), 1–11. https://doi.org/10.3389/fpsyg.2016.00690

Yarrow, L. J., Morgan, G. A., Jennings, K. D., Harmon, R. J., & Gaiter, J. L. (1982). Infants' persistence at tasks: Relationship to cognitive functioning and early experience. *Infant Behavior and Development, 5*(2–4), 131–141. https://doi.org/10.1016/S0163-6383(82)80023-4

Zumbo, B.D. (1999). A handbook on the theory and methods of differential item functioning (DIF).