The Unusual Box Test: A Non-Verbal, Non-Representational Divergent Thinking Test for Toddlers

Elena Hoicka[1,2], Simone Bijvoet-van den Berg[1], Tiffany Kerr[1], and Megan Carberry[1]

[1]Psychology, School of Natural Sciences, University of Stirling, Stirling, FK9 4LA, United Kingdom

[2]Psychology Department, University of Sheffield, Western Bank, Sheffield, S10 2TN, United Kingdom

elena.hoicka@stir.ac.uk

Abstract

Standard divergent thinking tests, e.g., the Wallach-Kogan Tests, and the Thinking Creatively in Action and Movement test, have verbal, representational, and imitative requirements limiting their use for children under 3 years. We present a new non-verbal, non-representational divergent thinking test that shows validity in relation to other standardized tests in 3- and 4-year-olds, and which is also reliable for use with toddlers as young as 19 months. This research is of value in understanding the early emergence of creativity. It could also aid research into Artificial Intelligence and robotics.

Introduction

A defining feature of humans is that we are able to adapt (e.g., Kirton, 1989). This is critical, as if we always imitate what has already been done, then we would likely not survive when changes hit our environment, such as climate change, or the collapse of the economy. Thus it is crucial that we can come up with novel, creative, ideas. This research topic is of interest to government, industry, education, and the arts (e.g., Department for Business, Innovation and Skills, 2009; Learning and Teaching Scotland, 2010), and yet has been little studied in early development. The advantages to researching creativity in toddlers are numerous. First, such research would allow us to examine the emergence of creativity, which could help answer questions as to how creativity originates. A knock-on effect of learning about the emergence of creativity in toddlers is that information about how creativity first develops, and which factors affect it, can be used to inform research in artificial intelligence and robotics on more effective strategies for developing creativity computationally. Finally, methods that test creativity earlier in development could be useful for testing robots as tests that work with younger children have fewer cognitive and linguistic demands.

Standardized tests of creativity exist which can be used with children, such as the Wallach-Kogan tests of creativity (1965), and the Torrance Creativity in Action and Movement test (TCAM, Torrance, 1981). The Wallach-Kogan tests require exclusively verbal responses, asking questions such as "Name all round things you can think of". Thus this test is too verbal for toddlers. The TCAM was created to rule out the requirement to answer verbally. It does so by modelling actions for children, e.g., two different ways to put a cup in a bin, such as hit the cup off your hand, or throw it overhand. Children are encouraged to copy the actions, and then to come up with different ways to do the action. This test works well with children as young as 3 years (Torrance, 1981). However the test creates two new limitations making it difficult for children under 3 years to participate. First, some of the subtests require pretense, introducing representational demands. Second, by initially having children copy actions, children must then inhibit this copying response to be original. Inhibitory control is difficult for 2-year-olds (e.g., Rennie, Bull, & Diamond, 2004), and in a pilot study in our own lab, we found floor effects when testing 2-year-olds as they simply copied actions we showed them, rarely inventing their own actions. Yet children as young as 6 months are shown to explore and acknowledge the affordances of different objects and surfaces (e.g., Bourgeois, et al., 2005). This suggests that infants and toddlers have the capacity to be creative through action and object exploration.

We have developed the Unusual Box test, which exploits children's exploration of objects, to measure divergent thinking. The test consists of a colorful box with a hole, stairs, ledges, and strings, along with novel objects (see Figure 1). Children are encouraged to play with the box, with no prior demonstration, and divergent thinking is measured through the number of action/box area combinations toddlers produce. Thus the task is non-verbal, non-representational, and non-imitative. Experiment 1 will assess the validity of the Unusual box test by comparing 3- and 4-year-olds' divergent thinking scores on subtests of the Wallach & Kogan tests (1965), the TCAM (Torrance, 1981), and the Unusual Box test. Experiment 2 will assess test-retest reliability of the Unusual Box test with 2-year-olds. Experiment 3 will assess the test-retest reliability of the Unusual Box test with 19- to 23-month-olds, and its relation to their level of motor development.

Experiment 1

The purpose of Experiment 1 was to determine whether the Unusual Box test is a valid measure of divergent thinking. To do so, we compared 3- and 4-year-olds' fluency and originality scores on the Unusual Box test to the fluency and originality scores on three existing divergent thinking tests: the Instances and Pattern Meanings subtests of the Wallach-Kogan (1965), and the TCAM (Torrance, 1981).

Method

Participants
Twenty-four 3- and 4-year-olds participated (13 males, *mean* = 46 months; range 37-57 months; *SD* = 6 months).

Materials
The Unusual Box is a wooden box (33.6x18x14.4cm). It contains the following features (see Fig. 1): (1) Ledges; three small blocks attached to an external wall of the box, and one shelf-like block. (2) Strings; 21 tie-wraps were tied across a wire so that they hang down the short side of the box. (3) Rings; seven closed tie-wraps in different sizes, attached to an external wall of the box to resemble rings. (4) Round hole; a hole (5.7cm in diameter) cut into the other short side of the box. (5) Rectangular room; a space of 10x5x8cm that can be reached via the round hole or the top of the box. (6) Stairs; two steps and a small edge on the top, covering two-thirds of the inside of the box. The stairs could be reached from the top of the box. The box was placed on a turntable (25cm diameter) to ensure that children could easily reach all sides of the box. Furthermore, five objects were used in the Unusual Box test, which were novel to the participants (see Fig. 1): a spiral-shaped egg holder, spatula, feather roller, kong rubber toy and hook.

Fig. 1 The Unusal Box and the five novel objects

Materials for the TCAM included paper cups, a small garbage bin and duct tape. Materials for the Pattern meanings subtest included cards with line drawings. The Unusual Box test and TCAM were recorded with two digital video camcorders (SONY Handycam) and tripods. The Instances and Pattern Meanings subtests were recorded with an Olympus MP3 recorder.

Design
A within-subjects design was used. All children were tested on the Unusual Box test, Instances, Pattern Meanings, and the TCAM. The order of tests was counterbalanced between children. Instances and Pattern Meanings were conducted together. For the Unusual Box test, the order of objects given to children was counterbalanced.

Procedure
For the Unusual Box test the experimenter placed the Unusual Box on top of the turntable. The experimenter highlighted each part of the box in the following order: ledges, strings, rings, round hole, rectangular room, and stairs. The experimenter turned the box while explaining so that the specific features were directly in front of the children. The children were given a chance to turn the box as well. Next, the children were told that s/he could play with the box together with another toy, until the experimenter instructed that s/he should stop. The children were then given one of five objects. They were given 90 seconds to play with each object, after which the object was replaced by a new one.

For the Instances and Pattern Meanings subtests the instructions were used as described by Wallach and Kogan (1965). Three out of four items of the Instances subtest were used. The items were presented in the following order: "Name all round things you can think of", "Name all the things you can think of that will make noise", and "Name all the things you can think of that move on wheels". The fourth item was omitted as pilot testing showed children did not understand the question.

For the Pattern Meanings subtest children were asked to name as many things as possible that a line drawing could be. Only the first four of nine items were used as pilot testing showed children did not pay attention for more items.

The instructions for the TCAM are described by Torrance (1981). Three of the subtests of the TCAM were used as the fourth did not involve divergent thinking. In the subtest 'What might it be?' the child had to think of as many possible different uses for a paper cup as they could. Two examples, using the cup as a hat and driving it around like a car, were demonstrated before the child could have a turn. In 'How many ways?' the child was asked to move between two lines (duct-tape attached to the floor) in as many ways as possible. As examples, walking and crawling were modeled. The last subtest was 'What other ways?' The child was asked to put paper cups into a garbage bin in as many different ways as possible. Two examples given were "putting the cup on the palm of your hand and shoving it in with the other hand", and "throwing the cup in the bin while standing a meter away from the bin". There was no time limit to the responses of the child in any subtest.

Coding

For the Unusual Box test, each trial started from the moment the child took the novel toy from the experimenter, and lasted 90 seconds. For each child two different types of scores were calculated: a fluency and originality score. The fluency score consisted of the number of different actions that the child performed for all trials combined (5 x 90 seconds). Actions were recorded on two features: the action performed (e.g. jump, hit, place) and the part of the box used during the action (e.g., ledges, hole). Thus a child would receive a score of 1 for repeatedly hitting the ledge, but a score of 2 for hitting the ledge and the stairs, as two locations were explored. Likewise a score of 2 would be received for hitting and sweeping the ledge as two actions were explored. Performance of the same action with different objects or the hand was given a score of 1. Inter-rater agreement for 20% of the videos was good ($k = 0.81$).

For originality scores, first, an originality index was created. Summing across children in Experiments 1 and 2, actions that were performed by fewer than 5% of children received a score of 3; actions performed by 5-20% of children received a score of 2; actions performed by 20-50% of children received a score of 1; and actions performed by more than 50% of children received a score of 0. Next, a total originality score was calculated for each child by adding up the originality scores of all the actions that s/he had performed. An originality ratio score was also calculated by dividing the total originality score by the Fluency score.

Fluency scores for the Instances Task were calculated by counting the number of different appropriate answers that a child gave. For example, when asked to "name all the round things you can think of" a circle was coded as a correct answer, while a brick was coded as incorrect. Originality scores were given for correct answers which no other children reported. The Pattern Meanings Test was scored in the same way.

Fluency scores for the TCAM were calculated by counting the number of different correct actions. For the "What might it be?" subtest, correct answers included actions with the cups that involved placing the cup in unusual places or building something out of several cups. The "how many ways?" subtest was coded for the number of times a child moved in a different way. For the "What other ways" subtest, correct answers included dropping the cup into the bin from one of the child's body parts (e.g. knee drop, arm drop, head drop), making specific movements before throwing the cup into the bin or putting the cup into the bin accompanied by something else. Lists of possible answers for all three subtests are given by Torrance (1981).

Originality scores were calculated following the manual provided with the TCAM (Torrance, 1981). Each response in the manual corresponds with an originality score. This score is based, "primarily upon the statistical infrequency of the response in a normative sample of five-hundred children" (Torrance, 1981, p. 15). Each separate response was given an originality score between 0 and 4. All scores were added up to provide a total originality score.

Results

No effects of gender were found in any analyses.

Fluency

Table 1 shows the descriptive statistics of the fluency scores for each test. Age was positively correlated to Instances and Pattern Meanings. Therefore, further analyses were corrected for age.

The correlations between the test scores are also given in Table 1. The Unusual Box test was positively correlated to Instances and TCAM, but not to Pattern Meanings. In fact, Pattern Meanings scores were not significantly correlated to any of the other tests.

			Correlations			
	Mean	Range	Age	1[a]	2[a]	3[a]
1.Unusual Box test	24.00 (6.5)	8 – 34	0.18			
2.Instances	8.33 (4.1)	2 – 18	0.49*	.49*		
3.Pattern Meanings	5.8 (2.0)	3 – 10	0.44*	.34	.22	
4. TCAM	91.58 (11.2)	71 – 114	-0.32	.60*	.60*	.02

Table 1. Means and standards deviations (in brackets), and correlations of fluency and scores among all divergent thinking tests in Experiment 1.

[a]Partial correlations, corrected for age. * $p < .05$.

Originality

Table 2 shows correlations of 0.35 to 0.41 between the originality scores on the Unusual Box test and originality scores on the other divergent thinking measures. However this is largely predicted by Fluency. When looking at the ratio of total Originality divided by Fluency, no correlations are found.

Discussion

Experiment 1 demonstrates that the Unusual Box test captures divergent thinking as the fluency scores for the Unusual Box test correlated well with the fluency and originality scores on the Instances subtest and the TCAM. The Unusual Box test is thus a valid measure for capturing divergent thinking. The Pattern Meanings subtest did not correlate with any other measures, suggesting it is not suitable for use with 3-year-olds.

Originality scores on the Unusual Box test correlated well with originality scores on the other tests, however this was driven by Fluency. This confoundedness of fluency and originality scores is not uncommon (e.g. Clark &

Mirels, 1970). Thus increased fluency may lead toddlers to be more original, or vice versa.

	Mean	Range	Correlations 1[a]	2	3	4
1. Unusual Box	21.17 (9.0)	4 – 41				
2. Instances	2.96 (2.3)	0 – 8	.41*			
3. Pattern Meaning	2.04 (1.3)	0 – 5	.41*	.16		
4. TCAM	95.83 (9.6)	80 – 112	.35[†]	.22	.22	
5. UB Fluency	24.00 (6.5)	8 – 34	.88*	.46*	.41*	.50*
6. UB originality ratio	.87 (.24)	.39 – 1.42		.28	.29	.19

Table 2. Means and standard deviations (in brackets) and correlations of originality scores among all divergent thinking tests in Experiment 1. UB = Unusual Box
[a] Spearman's Rho instead of Pearson's *r*, *$p < .05$, †$p < .1$

Experiment 2

Experiment 2 investigated whether the Unusual Box test is suitable for children younger than three years. We investigated the test's reliability with a test-retest design.

Method

Participants

Sixteen two-year-olds participated (7 males, mean age = 28 months, range 24-32 months, SD = 3 months).

Materials

Same as the Unusual Box test materials for Experiment 1.

Design

A within-subjects design was used. All children played with the Unusual Box test twice, two weeks apart. Counterbalancing of objects was the same as Experiment 1. Children received different object orders at each time of testing.

Procedure

Same as the Unusual Box test procedure in Experiment 1.

Coding

Same as the Unusual Box test coding in Experiment 1.

Results

No effects of gender or age were found in any analyses.

Fluency

The average fluency scores were 19.3 actions (SD = 5.9, *range* 10-32) at Time 1, and 20.1 at Time 2 (SD = 6.2, range 12-36). No differences in scores were found between Time 1 and 2 (paired-sample $t(15) = 0.856$, $p = 0.406$), indicating that children obtained similar scores on both assessments. A strong positive correlation was found between the scores of the two assessments (Pearson's r = 0.58, $p = 0.019$).

Originality

The average originality-ratio scores were 0.76 (SD = 0.21, *range* 0.39-1.13) at Time 1, and 0.74 at Time 2 (SD = 0.25, *range* 0.33-1.19). No differences in scores were found between Time 1 and 2 (paired-samples $t(15) = 0.037$, p = 0.971), indicating that children obtained similar scores on both assessments. A strong positive correlation was found between the scores of the two assessments (Pearson's r = 0.805, $p < 0.001$).

Discussion

Experiment 2 demonstrates that the Unusual Box test is reliable for use with 2-year-olds. This is thus the first non-verbal, non-representation, non-imitative divergent thinking test that can be used with children under 3 years.

Experiment 3

The aim of this study is to discover whether the Unusual Box test can be used with toddlers under 2 years. An additional aim is to examine whether motor development is linked to toddlers' divergent thinking scores on the Unusual Box test. This will mark the earliest time point at which we can currently look at potential influences on creativity, such as parenting style, and other cognitive abilities.

Method

Participants

Twelve 19- to 23-month-olds participated (7 males, *mean* = 21.5 months, SD = 1.6 months). Eight of the children have to date completed the 2-week follow-up with the Unusual Box (6 males, *mean* = 22.5 months, *range* = 21-23 months; SD = 0.75 months)

Materials

Same as Experiment 2. The Child Development Review (CDR; Ireton, 1990) was used to assess toddlers' fine and gross motor-skill development through a parent check list.

Design

Same as Experiment 2. Children's fine and gross motor development served as additional independent variables.

Procedure
Same as Experiment 2, with the addition of asking parents questions from the CDR after the toddler played with the Unusual Box.

Coding
Same as Experiment 2. Fine and gross motor skill ages were determined from the parent reports.

Results
No effects of gender or age were found in any analyses. The mean scores for all 12 participants on the Unusual Box test at Time 1 was 16.2 actions (SD = 5.6, *range* 8-26); and for the 8 participants who participated in both assessments, 15.3 actions (SD = 5.5, *range* 8-25) at Time 1, and 16.7 actions (SD = 3.5, *range* 13-22) at Time 2. No differences in scores were found between the eight children at Times 1 and 2 (paired-samples $t(7) = 1.27$, $p = 0.244$), indicating that children obtained similar scores on both assessments. A strong positive correlation was found between the scores for the eight children at Times 1 and 2 (Pearson's r = 0.818, p = 0.013). Additionally, fine motor skills age (Pearson's r = 0.629, p = 0.029) and gross motor skills age (Pearson's r = 0.667, p = 0.018) both had large correlations with the Unusual Box fluency score at Time 1 across all 12 participants. Since gross and fine motor skills ages were highly correlated (Pearson's r = 0.725, p = 0.008), it is difficult to determine whether one or both of these factors are more closely related to divergent thinking scores.

Discussion
Experiment 3 demonstrated that the Unusual Box test is reliable for use with toddlers as young as 21 months. It was also found that toddlers' fine and gross motor skills were highly correlated with their divergent thinking scores as young as 19 months. We will collect more data with toddlers as young as 12 months to further examine how young we can use the test, and to further explore the role of motor development. We will also calculate originality scores once the full sample has been tested.

General Discussion

Experiment 1 demonstrated that the Unusual Box test is a valid tool for assessing divergent thinking as it correlates well with other standardized tests of divergent thinking (Instances, TCAM) in 3- and 4-year-olds. Experiments 2 and 3 demonstrated that the Unusual Box test is reliable for use with children as young as 21 months. This test thus provides a tool to examine the emergence of divergent thinking at an earlier age than has been possible to date. Experiment 3 also found that fine and gross motor skills were highly correlated with divergent thinking frequency scores, while age was not, for toddlers as young as 19 months.

It is important to note that the correlations between the different divergent thinking tests in Experiment 1 were moderate, but not high. This was not only the case for the Unusual Box test in relation to the other tests, but was also the case between the other tests as well. We argue that the moderate correlations support the notion that these tests capture divergent thinking. However differences in the format of these tests, for example, the extent that they involve verbal, representational, and visual information, likely introduce variance in the scores. Thus children who are more verbal may do better on the Instances subtest, or some children may be more verbally creative, while others are more physically creative. Future research should examine how different cognitive, communicative, and physical abilities interact with divergent thinking skills, leading to individual differences in creative outputs in different domains.

An interesting question results from Experiment 3. Is it the case that advanced fine and gross motor skills allow toddlers to explore objects in more diverse ways, or is it possible that toddlers who have higher divergent thinking abilities develop their fine and gross motor skills more quickly because they are motivated to explore? Longitudinal research should address this question. In doing so, we may also develop a better idea as to whether divergent thinking abilities and individual differences are to some extent innate, and can later demonstrate themselves once children develop their motor, language, and representation skills. This claim would suggest that computer programs should have intrinsic motivational properties to express divergent thinking, which may be expressed as autonomy, self-motivation, self-direction, independence, or freedom (e.g., al-Rifaie, Bishop, & Caines, 2012; Jordanous, 20120; Saunders, 2012). Conversely, it may be that the development of these skills give toddlers an ability to think divergently. This would suggest that divergent thinking could emerge as a property of developing motor, language, and/or representational skills, lending support to current embodied approaches to creativity in AI (e.g., Saunders, et al., 2010).

The development of this test is important as it will allow us to examine which factors affect divergent thinking early on. One potential factor is parenting. Parents have been shown to directly scaffold infants' and toddlers' understanding of two types of creative acts – joking and pretending – through cues such as language, actions, and laughter (e.g., Hoicka & Gattis, 2012; Hoicka, Jutsum & Gattis, 2008; Lillard & Witherington, 2004; Mireault, et al., 2012). Thus the way a parent interacts with their child may be key to individual differences and learning of creativity.

Another area to explore is social learning's role in the early development of creativity. While Bonawitz, et al. (2011) suggest that imitation reduces toddlers' drive to explore, we posit that imitating abstract concepts, rather than direct exemplars, could encourage creativity more generally. For instance, toddlers come up with their own mislabeling jokes (e.g., calling a cup, “goojooboojoo”) after being exposed to other mislabeing jokes (Hoicka & Akhtar, 2011). Work in progress (Bijvoet-van den Berg & Hoicka, in prep; Bijvoet-van den Berg, Liszkowski & Hoicka, in prep.) suggests 3-year-olds create novel acts of pretense through object substitution, and novel iconic gestures, after copying similar acts. Future research can thus use the Unusual Box test to determine whether toddlers can learn to explore more broadly through social learning.

The main advantages of the Unusual Box test are that it is non-verbal, non-representational, and non-imitative, making it ideal for use with toddlers. Due to its nature, it may also be useful with older children and adults with communicative disorders, such as Autism, Speech and Language Impairment, and deaf children born to non-signing parents, as the verbal demands are very limited. This type of task could also be of use in examining the phylogeny of creativity as a variation of this task could be used with primates. Finally, this paradigm is relevant for use in Artificial Intelligence and robotics in two ways. First, it can directly provide a tool to examine object exploration in robots. Second, by further examining physical, social, cognitive, emotional, and other factors that affect creativity in early development, we can better understand how creativity emerges, allowing for more sophisticated computational models of creativity to be developed.

Acknowledgments

This research was funded by a PhD studentship from Psychology, School of Natural Science, University of Stirling, to Simone Bijvoet-van den Berg. We thank children for participating in our research, and Stirling Council and parents for allowing children to participate in our research. We also thank Joanne Kirkwood and Rachael Mowat for help with data collection, and Claire Beattie, Angela Boyle, and Beatrice Bretherton for help with data coding.

References

al-Rifaie, M. M., Bishop, J. M., & Caines, S. (2012). Creativity and autonomy in swarm intelligence systems. *Cognitive Computation* 4: 320-331.

Bijvoet-van den Berg, S., & Hoicka, E. (in prep). Preschoolers understand intentions to produce object-substitution, and produce their own novel acts of object substitution.

Bijvoet-van den Berg, S., Liszkowski, U., & Hoicka, E. (in prep). Preschoolers communicate using novel iconic gestures.

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: 322-330.

Bourgeois, K. S., Khawar, A. W., Neal, S. A., & Lockman, J. J. (2005). Infant manual exploration of objects, surfaces, and their interrelations *Infancy,* 8: 233-252.

Clark, P. M., & Mirels, H. L. (1970). Fluency as a pervasive element in the measurement of creativity. *Journal of Educational Measurement* 7: 83-86.

Department for Business, Innovation and Skills. (2009). *Thinking Business in Policy: An Interim Report.* HM Government.

Hoicka, E., & Akhtar, N. (2011). Preschoolers joke with jokers, but correct foreigners. *Developmental Science* 14: 848-858.

Hoicka, E., & Gattis, M. (2012). Acoustic differences between humorous and sincere communicative intentions. *British Journal of Developmental Psychology* 30: 531-549.

Hoicka, E., Jutsum, S., & Gattis, M. (2008). Humor, abstraction, and disbelief. *Cognitive Science: A Multidisciplinary Journal* 32: 985-1002.

Ireton, H. (1990). Developmental screening measures. In J. H. Johnson, & J. Goldman (Eds.) *Developmental Assessment in Clinical Child Psychology: A Handbook* (pp. 78-99). Elmsford, New York: Pergamon Press.

Kirton, M. J. (1989). *Adaptors and Innovators: Styles of Creativity and Problem Solving.* London: Routledge.

Jordanous, A. A Standardised Procedure for Evaluating Creative Systems: Computational Creativity Evaluation Based on What it is to be Creative.*Cognitive Computation* 4: 246-279.

Learning and Teaching Scotland. (2010). *Pre-birth to Three: Positive Outcomes for Scotland's Childen and Families.*Scottish Government.

Lillard, A. S., & Witherington, D. C. (2004). Mothers' behavior modifications during pretense and their possible signal value for toddlers. *Developmental Psychology* 40: 95-113.

Mireault, G., Poutre, M., Sargent-Hier, M., Dias, C., Perdue, B., & Myrick, A. (2012). Humour perception and creation between parents and 3- to 6-month-old infants. *Infant and Child Development,* 21: 338-347.

Rennie, D. A., Bull, R., & Diamond, A. (2004). Executive functioning in preschoolers: Reducing the inhibitory demands of the dimensional change card sort task. *Developmental neuropsychology* 26: 423-443.

Saunders, R. (2012). Towards autonomous creative systems: A computational approach. *Cognitive Computation* 4: 216-225.

Saunders R, Gemeinboeck P, Lombard A, Bourke D, Kocabali B. Curious whispers: an embodied artificial creative system. In: International conference on computational creativity 2010, 7–9 January 2010. Lisbon; 2010.

Torrance, E. P. (1981). *Thinking Creatively in Action and Movement (TCAM).* Bensenville, IL: Scholastic Testing Service, Inc.

Wallach, M. A., & Kogan, N. (1965). *Modes of Thinking in Young Children.* Holt, Rinehart and Winston: New York.