*Dunn, A., *Qiao, A., *Johnson, M., & Kunda, M. (2021). "Measuring more to learn more from the block design test: A literature review." Proceedings of the 43rd Annual Meeting of the Cognitive Science Society. p. 611-617. *Co-first authors.

Measuring More to Learn More From the Block Design Test: A Literature Review

Avery Dunn[*1] **(dunnavery@wustl.edu), Alice Qiao**[*] **(alice.qiao@vanderbilt.edu)**
Maya Johnson[*2] **(mjohnson5@brynmawr.edu), Maithilee Kunda (mkunda@vanderbilt.edu)**
[*] These authors contributed equally to this work.
Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37235 United States
(Present affiliations: [1]Washington University, St. Louis, MO, USA; [2]Bryn Mawr College, Bryn Mawr, PA, USA)

Abstract

The block design test (BDT), in which a person has to recreate a visual design using colored blocks, is notable among cognitive assessments because it makes so much of a person's problem-solving strategy "visible" through their ongoing manual actions. While, for decades, numerous pockets of research on the BDT have identified certain behavioral variables as being important for certain cognitive or neurological hypotheses, there is no unifying framework for bringing together this spread of variables and hypotheses. In this paper, we identify 25 independent and dependent variables that have been examined as part of published BDT studies across many areas of cognitive science and present a sample of the research on each one. We also suggest variables of interest for future BDT research, especially as made possible with the advent of advanced recording technologies like wearable eye trackers.

Keywords: cognitive strategy; neuropsychological testing; spatial skills; visuospatial reasoning.

Introduction

One hundred and one years ago, Samuel Kohs published the manual for his Block Design Test (BDT), in which a person has to rearrange colored blocks to recreate visual designs of increasing difficulty (Kohs, 1920). The BDT has seen incredibly prolific use for the past century, both as a standalone test and as a subtest on larger batteries like the Wechsler intelligence scales.

The BDT is routinely used as a measure of visuospatial ability in research, educational, and clinical settings, with performance scored as a function of per-item accuracies and response times. However, the uniquely physical nature of the test—i.e., the test-taker is continuously manipulating blocks to construct their answer—has led to numerous studies of the *process* by which people solve BDT items. In particular, the BDT offers many opportunities to add independent and/or dependent variables to the mix that can help reveal aspects of people's otherwise unobservable cognitive strategies.

In fact, Kohs originally thought of his test as a general intelligence test precisely because it offers such a rich problem-solving environment to the test-taker (Kohs, 1920, p. 370):

> "If 'intelligence' involves the following mental operations; analyzing, combining, comparing, deliberating, completing, discriminating, judging, criticising and deciding, then the block-design tests may, with justice, be said to call upon the functioning of intelligence and to that extent they are a measure of that mental capacity."

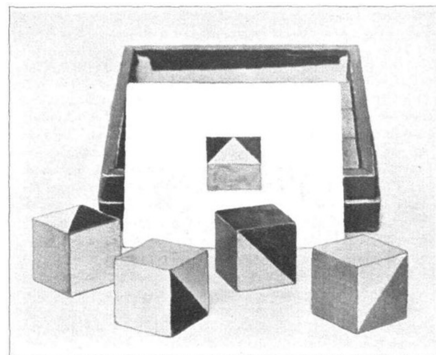

Figure 1: An illustration of the block design test (BDT) from Kohs' original paper (Kohs, 1920).

Interestingly, Kohs originally proposed that the BDT be scored based on accuracy, time, and number of moves, with a composite score that integrated all three measures. The scoring method was later revised, for ease of test observation and scoring, to use just accuracy and time (Hutt, 1932).

While most BDT administrations use this simpler scoring scheme, a wealth of studies over the decades have examined more nuanced measures of BDT performance in different populations, including healthy adults, people with brain injuries, neurodiverse populations, and the very young or old. These "extra" measurements range from error types, to block ordering patterns, to gaze sequences. In addition, the past two decades have seen advancements in technology, such as gaze-tracking devices and computerized versions of the BDT, that allow researchers to increase the type, density, and precision of behavioral measurements that can be collected.

Having this broader behavioral picture of a person's BDT performance can help researchers extract more inferences from BDT results, as well as identify gaps in knowledge and future research paths. In this paper, our contributions include:

- We identify 25 different independent and dependent variables that have been examined as part of published BDT studies across many areas of cognitive science.
- In addition to explaining each variable, we present a sample of the research on each one.
- Finally, we suggest variables of interest for future BDT research, especially as made possible with the advent of advanced recording technologies like wearable eye trackers.

Method

We conducted a survey of the BDT literature to identify studies that adopted any kind of "extra" measures into the test administration procedure, including the addition of independent variables (e.g., features of BDT designs given to test-takers) as well as dependent variables (e.g., error patterns, gaze, etc.).

We started by searching for BDT papers using Google Scholar. Search terms included "block design," "block design task," "block design study," etc. We also combined these terms with populations that cropped up in initial papers, such as "dyslexia," "brain injury," "autism," etc., as well as specific variables, such as "errors," "broken configuration," etc.

After finding a paper, we first labeled it as a review versus an empirical study, and we focused our remaining examination just on papers that presented original empirical data. We eliminated papers that used just the standard BDT test with standard scoring based only on accuracy and response time. We then identified what "extra" independent and/or dependent variables each paper reported. Finally, for each paper we found, we expanded our search for additional papers to include looking at both forward and backward references.

Our survey yielded many published papers meeting the above criteria. Due to space constraints, and the ongoing nature of our survey, we do not include the full list of papers here, but we do present the full list of BDT variables that we identified through this process. In the following sections, we first describe independent variables and then dependent variables. A more detailed list of papers, along with demographics of each study and brief highlights of results, can be found in online supplementary material at:

`https://doi.org/10.6084/m9.figshare.14560593.v1`

Figure 1 shows the number of papers in our survey that studied each of the independent and dependent variables on the BDT. Many papers studied two or more variables.

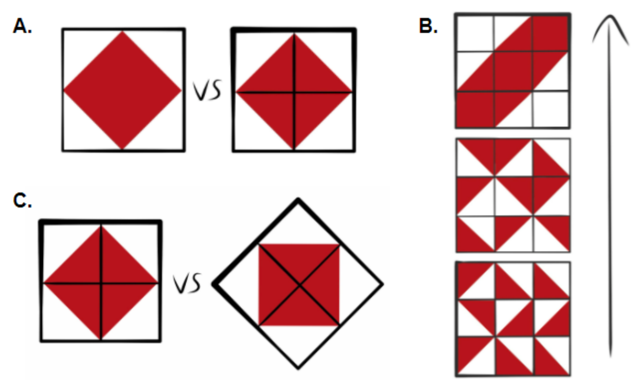

Figure 2: Block Design Test Independent Variables. **(a)** Whole vs segmented design. **(b)** Perceptual coherence (low to high). **(c)** Different design orientations.

Independent Variables

1. Whole vs Segmented designs. A multitude of studies have been conducted to test whether the segmentation of the design has an effect on performance (Figure 2a). Stewart et al. (2009) measured whether those with high and low autism quotients performed better or worse on segmented or unsegmented (whole) designs. The findings were that groups that had a high autism quotient did better on the unsegmented designs but groups performed similarly on both types of designs. Across the many studies, the general trend is in the direction that neurotypical and autistic individuals perform better on segmented designs, but autistic individuals perform better than neurotypical individuals on unsegmented designs.

2. Perceptual Coherence. Perceptual coherence is a measure of global coherence (i.e. there are discernible patterns or shapes within the design) versus local coherence (i.e. the

Table 1: Number of papers that studied each variable. Full list: `https://doi.org/10.6084/m9.figshare.14560593.v1`

	Variable Name	Papers
Independent	Perceptual Coherence	15
	Whole vs Segmented designs	11
	Spatial layout of task (dominant side vs non-dominant)	1
	Spatial layout of task (cost of gaze shifts and motor actions)	1
	Different orientations of the designs	1
Dependent Variables	Variables Related to Block Placement	24
	Trial-and-error block placement attempts	13
	Overall Cell-ordering pattern	8
	Contiguous Cell ordering	5
	Trial-and-error block changes	4
	Starting position	4
	Block rotations in hand	4
	Time per block placement	3
	Orientation of selected blocks	3

	Variable Name	Papers
Dependent Variables	Variables Related to Errors	16
	Broken Configuration	13
	Rotated design	5
	Single-block rotations	5
	Rearrangement of Gestalts	3
	Closing-in	2
	Odd angles	1
	Non-squared blocks	1
	Stacking	1
	Variables Related to Gaze	8
	Frequency of Consulting the Pattern	7
	Gaze Location	2
	Other Variables	4
	Within-Test Practice Effects	3
	Motor Actions	1

blocks are randomly placed with no discernible shapes). For example, the bottom-most design in Figure 2b has low perceptual coherence, while the top one has high perceptual coherence. This analysis of local versus global coherence has been conducted in a plethora of studies. While findings vary across studies, the conclusions trend in the direction that low perceptual cohesiveness (local) is correlated with better performance whereas high perceptual cohesiveness (global) is correlated with worse performance.

3. Different Orientations of Designs. BDTs have been conducted where the participant must complete a design that is both oriented squarely in front of them and at a 45 degree angle (Figure 2c). A popular study conducted by Frith et al. (1995) employed this type of test, where they oriented the design at a different angle for the participant to complete. The effect on performance and timing can be measured, compared, and analyzed for each differently oriented design.

4. Spatial layout of task (cost of gaze shifts and motor actions). The spatial layout has an effect on where the participant maintains a focus. The layout of the task can have an effect on the cost of gaze shifts, as measured by Burggraaf et al. (2016), who used a pencil-and-paper version of the block design to measure gaze shifts.

5. Spatial layout of task (dominant side versus non-dominant side). This is a measure of whether the setup or layout of the BDT, specifically whether the block bank is placed on the right or left side of the participant, and whether picking up blocks with the dominant or non-dominant hand has an effect on timing. A measure on handedness in the BDT was conducted in a study by Begum et al. (2017). There was no significant difference found among the results but a slight favor in right-handed scores.

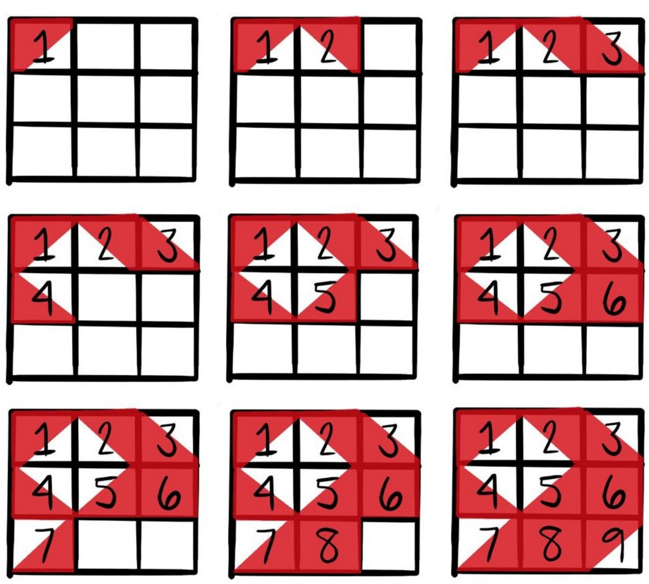

Figure 3: Sample Cell Ordering Pattern.

Dependent Variables

Variables Related to Block Placement

6. Overall Cell Ordering Pattern. This measure describes the order in which the participant places the blocks, one example of which is shown in Figure 3. The ordering pattern is commonly studied, and has been measured in individuals with autism, typically developing children with dyslexia, children with nonverbal learning disability, and different age groups. This measure was first pointed out by Jones and Torgesen (1981), where they used this variable alongside other variables to compare strategy between older and younger children. In Rozencwajg and Corroyer (2001), this variable was used alongside other variables to identify different strategies (global, analytic, and synthetic) between different age groups.

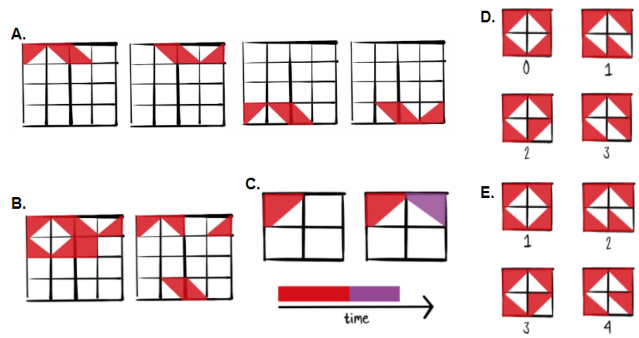

Figure 4: Block Placements. **(a)** Construction starting position. **(b)** Contiguous vs non-contiguous cell ordering. **(c)** Time per block placement during construction. **(d)** An example of block changes and numbering. **(e)** An example of block placement and numbering.

7. Starting Position. Studies recorded whether subjects started on the left/right, top/bottom, and center to determine if participants had a preference for a starting side (Figure 4a). This variable has been measured in several studies. It was first pointed out by Jones and Torgesen (1981), where it was used to compare strategy across different divisions of the participants, including age. They found older children did not consistently use different strategies than younger children (including where they started). Later, Akshoomoff et al. (1989) recorded the starting point (left/right) to measure differences between alcoholic and unilateral brain-damaged patients.

8. Contiguous Cell Ordering. This variable measures if participants place blocks in adjacent sections or in disconnected sections (Figure 4b). This variable is not commonly measured and was first identified by Dirks (1982) to measure differences between children with and without experiences with a commercial game similar to the BDT.

9. Time per Block Placement. This variable measures the amount of time needed to place a block correctly in a cell (Figure 4c). It is not commonly measured. In Schorr et al. (1982), they measured a version of this variable; they tested the time to place a block as a function of solid vs two-colored blocks, empty (placing first block) vs filled construction space (placing last block), and number of interior edge cues.

10. Trial-and-error Block Changes. The number of times a participant tries to fix errors after the first placement (i.e., the first placement does not count as a change) (Figure 4d). This can be measured per block or per cell. The number of trial-and-error block changes is equal to the number of trial-and-error block placement attempts minus one. This variable is not commonly measured. It was first used by Joy et al. (2001), where they counted the number of changes ("the number of times a block was rotated in place or moved to a different location after having been placed") made by healthy older adults. In Hoffman et al. (2003), they identified attempts to fix errors during and after building process, finding that children complete more fixes and checking of designs during the process.

11. Trial-and-error Block Placement Attempts. Participants tend to follow a pattern of manipulating a block, placing it down, checking if it is correct, and pick it back up to re-manipulate it until it is correct. This variable measures the number of attempts used to place a block correctly (Figure 4e). It can be measured as the number of attempts per block or the number of attempts per cell. This variable is very commonly measured, and has been measured in many groups including individuals with autism, neurotypical individuals, learning disabled individuals, and individuals with asperger syndrome. It was first measured by Jones and Torgesen (1981). Later, Troyer et al. (1994) measured multi-block (two blocks correct with respect to each other but not the overall design) and single-block errors in healthy older adults, both in the final design and those self-corrected during construction. They found that these were very common.

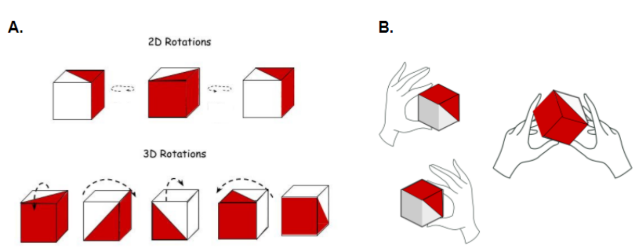

Figure 5: Rotation and Other Motor Actions. **(a)** 2D and 3D rotations. **(b)** Block manipulation with different hands.

12. Block rotations in hand. This variable is very commonly measured. Some studies measure the number of manipulations a participant makes for a block. Other studies measure whether a participant did or did not rotate the block before placement. This can be further broken down into 2D and 3D rotations. A 3D rotation is defined as rotating to a new face of the block. A 2D rotation is defined as changing the orientation of one face of the block. Examples of both types of rotations can be seen in Figure 5a. This variable was first measured by Salthouse (1987), where they used an adapted computerized version of the block design test and recorded the number of block manipulations as a measure of efficiency. Another example of this measure is in Rozencwajg (1991), where block rotations were used to help calculate anticipation scores; high anticipation implies the individual rotated the block in their hand before placement.

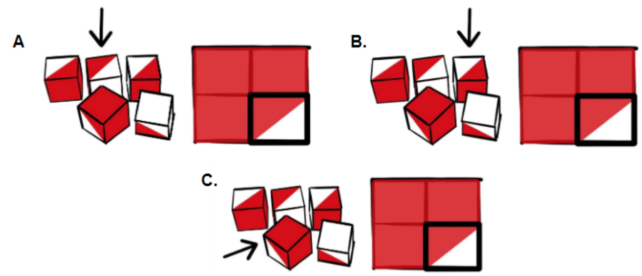

Figure 6: Block Orientation Selection in Block Bank. **(a)** Block selected in the correct orientation from the block bank. **(b)** Block selected that requires a 2D rotation. **(c)** Block selected that requires a 3D rotation.

13. Orientation of Selected Blocks. This measures whether or not a participant picks a block that is already in the correct orientation from the block bank (Figure 6). In other words, the participant either 1) has a specific cell and block orientation in mind and chooses a block with that needed orientation, or 2) picks up a block form the block bank and then rotates it to the correct orientation. This variable is relatively common to measure. A version of this variable was first pointed out by Schorr et al. (1982). Subjects were asked to identify which block face from a set of blocks matched the pattern for a marked target cell. This variable contributed to their analysis of strategy and cognitive deficit in completing the block design test. Later, this variable was similarly measured by Hoffman et al. (2003). They found children with William's syndrome struggle with spatial representations and struggled to match specific block orientations to the correct match in the design.

Variables Related to Errors

14. Broken Configuration. This is an error where a block is placed outside of the area of the design (Figure 7e). This error is common to measure, especially in people with brain injuries, neurodiverse populations, and older adults. It was first measured by Ben-Yishay et al. (1971) in people with brain damage, where they found that people with left hemisphere injuries made more of these errors. However, in most future papers, such as Akshoomoff et al. (1989), people with right hemisphere injuries were found to make more broken configuration errors. In Joy et al. (2001), this error was measured in older adults, where it was found to increase with age, but was also demonstrative of a trial-and-error strategy.

15. Rotated Design. This is an error where the whole design is rotated on an axis (Figure 7c). It is fairly common to measure, and is studied most in people with brain injuries, older adults, or in the general population to determine broad strategy. It is often grouped with broken configuration errors. Schatz et al. (2000) studied rotation errors in children with right or left hemisphere brain injuries, and found that those with left hemisphere injuries made more rotation errors. Kramer et al. (1991) found that people who made rotation errors or broken configuration errors focused more on local pattern similarities than people who did not.

16. Odd Angles. This is an error where a block is placed at an angle to previously placed blocks (Figure 7b). It is very uncommon to measure, appearing in only one identified paper. In this paper, odd angle errors were measured in 177 older adults and were found to increase with age. They also found that participants who used a trial-and-error approach to solving made more odd angle errors (Joy et al., 2001).

17. Non-squared Blocks. This is an error where a block is placed flush with previous blocks but not aligned at the corners (Figure 7d). It is very uncommon to measure, appearing in only one identified paper. In this paper, non-squared block errors were measured in 177 older adults and were found to have no significant change with age (Joy et al., 2001).

18. Stacking. This is an error where blocks are placed on top of each other vertically (Figure 7a). It is very uncommon to measure, appearing in only one identified paper. In this paper, stacking errors were measured in 42 neurologically impaired and 225 healthy older adults. Stacking errors had a medium effect size, but was not significantly significant (Paolo & Ryan, 1994).

19. Closing-in. Also called stimulus boundedness, this is an error where the blocks are placed on top of or next to the edge of the design (Figure 7f). It is fairly uncommon to measure, but is usually measured in older and neurologically impared populations. In Paolo and Ryan (1994), closing-in was measured in 42 neurologically impaired and 225 healthy older adults. Closing-in had a medium effect size, but was not significantly significant. In Troyer et al. (1994), it was observed to occur once in 145 healthy older adults.

20. Single-block Rotations. This is an error where a single block is incorrectly rotated (Figure 7g). It is a common error, especially during the construction process, but is fairly uncommon to study. It has been measured in neurodiverse or culturally distinct populations, such as Indigenous communities, and can be counted in the final design or during construction. Vasterling et al. (2000) found that people with PTSD made significantly more single block rotation errors during construction, even though they performed similarly to the control group on other measures. Pontius (1993) measured single block rotation errors in urban Indonesians and the Dani and Asmat people and found that the Indigenous groups made more of these errors.

21. Rearrangement of Gestalts. This is an error where recognizable shapes in the design (gestalts) are reordered, flipped, inverted, or otherwise rearranged (Figure 7h). It is not commonly studied, appearing in only one author's work. Pontius (1993, 1995, 1997) all compare gestalt rearrangement in a few different Indigenous and non-Indigenous populations. They conclude that Indigenous individuals perform more rearrangement of gestalts because of environmental necessities and cultural differences in thinking.

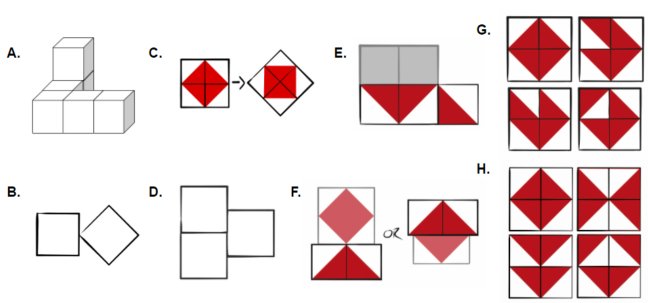

Figure 7: Types of Errors on the Block Design Test. **(a)** Stacking. **(b)** Odd angles. **(c)** Design rotation. **(d)** Non-squared blocks. **(e)** Broken configuration. **(f)** Closing-in. **(g)** Single block rotation. **(g)** Gestalt rearrangement.

Variables Related to Gaze

22. Gaze Location. This variable records where on the table the person is looking (at the design, block bank, construction area, etc) (Figure 8a). It is very uncommon to study, partially because of the lack of technology to accurately measure it until recently. In Kunda et al. (2016), gaze location is included in their computational architecture of the block design test, and is proposed as a determinant of strategy. In Cha et al. (2020), they show the feasibility of combining overhead and gaze cameras to record aspects of performance, including gaze location, and also suggest gaze as a measure of differentiating participants' strategies and memory usage. Gaze has also been studied in detail in the context of non-BDT but similar block copying tasks (Hayhoe et al., 1998).

23. Frequency of Consulting the Pattern. This measures how often the participant looked at the target design during construction (Figure 8b). The frequency of consulting the pattern can be measured physically, using gaze tracking technology, or virtually, e.g., by having participants click to reveal the pattern in a computerized BDT. It has been used to look at individual differences in children and adults in the the general population, but has also been used to look at age-related

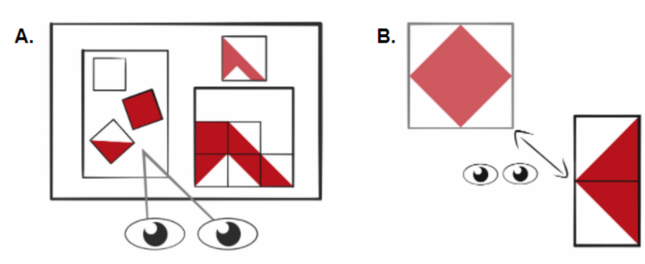

Figure 8: Gaze Patterns on the Block Design Test. **(a)** Gaze location. **(b)** Frequency of consulting the pattern.

differences. It is often combined with other variables to quantify different strategies. A majority of studies on this variable were carried out by the French researcher Rozencwajg and colleagues (Rozencwajg, 1991; Rozencwajg & Corroyer, 2001; Rozencwajg et al., 2005; Rozencwajg & Fenouillet, 2012; Fenouillet & Rozencwajg, 2015). The frequency of pattern consultation was first studied in Rozencwajg (1991), where they used it to help differentiate three main construction strategies. Hoffman et al. (2003), studied this in children with and without William's syndrome; children with William's syndrome looked at the design less frequently than the control group.

Other Variables

24. Motor Actions. This measures whether the participant manipulates (i.e. picks up, rotates, places down) a block using the left hand, right hand, or both hands (Figure 5b). A study on this variable was carried out by tracking the "leading hand" used to pick up and manipulate the blocks. This particular study, done by DeLuca et al. (1990), found that there was a shift towards the usage of the left as the leading hand.

25. Within-Test Practice Effect. This measures if participants perform better on items later in the BDT, often measured by comparing the time taken to complete earlier vs. later items. Miller et al. (2009) measured learning effects by varying the presentation order of the design based on perceptual cohesiveness and set size uncertainty. The findings were that varying the designs can help reduce the learning effect on score results.

Discussion and Future Work

Behavioral measurements can provide valuable insight into the cognitive strategies used by people completing the block design test (BDT). Two people could have very similar overall scores based on response time and accuracy but approach the task in very different ways; the simpler scoring scheme flattens these rich individual differences into looking like the same thing. In our review, we identified 25 different variables in published BDT studies that allow for a more nuanced understanding of people's task performance.

Of the independent variables, perceptual coherence (variable #2) and whole vs. segmented designs (variable #1) appear to have a more notable impact on performance than other variables. These variables are useful for manipulating the difficulty or complexity level of BDT items.

In terms of dependent variables, when measuring variables related to errors (excluding single-block rotations), many studies agree that a small number of errors is normal, but a high number of one error or combination of errors often correlates with the presence of a neurobehavioral difference (e.g., dyslexia or old age). Additionally, variables related to gaze have been increasingly studied as technology becomes more available and more integrated into the administration of the BDT (whether in computer-based BDT settings or using wearable eye-trackers or other cameras in physical BDT settings), and these variables may be particularly useful for understanding attention and memory on the BDT.

One issue facing this type of research is that there is a lack of standardization in the measurement methods for particular variables across different studies. For example, starting position has been categorized more broadly as left vs. right, but also with more detail as 1) left/right/center and 2) upper/lower, left/right/center. In addition, researchers or clinicians may aggregate two or more lower-level variables to create "higher-level" variables, such as strategy or internal representation of the block. For example, Rozencwajg and Corroyer (2001) integrated several dependent variables (overall cell ordering pattern, trial-and-error block placement attempts, and frequency of consulting the pattern) to define a small number of qualitatively different strategies.

In the future, it may be practical to implement standardized methods for the most commonly measured variables such as overall cell ordering pattern and broken configuration. Increased standardization would make it easier to compare results across studies. However, too much standardization may limit new ideas and discoveries, so it would be important for the community to balance the advantages of both approaches.

Recent work on other tests like the clock drawing test (Davis, Libon, Au, Pitman, & Penney, 2015) and on the BDT (Cha et al., 2020) has pointed out new roles that technology can play in advancing cognitive assessment methods. These roles include (Kunda, 2019): 1) advanced sensors (such as depth cameras, wearable eye trackers, etc.) or computer-based interfaces that can record rich streams of raw behavioral data; 2) pattern recognition algorithms that can be used to convert raw data (e.g., pixels) into meaningful behavioral measurements (e.g., block rotations); and 3) data mining algorithms that can be used to extract interesting higher-level patterns from sequences of behavioral measurements.

We close with examples of other variables not found in our review but that we expect would be of interest:

- The orientation of the blocks in the block bank, i.e., when the test-taker is presented with the set of blocks to use, do they take advantage of how the blocks are already oriented?
- Duration of first fixation (the time before construction begins), i.e., how long does the test-taker spend initially inspecting the design before starting to grab blocks?
- Adding one block vs. adding 2 or more during construction, i.e., to what extent does the test-taker use both hands? (This variable is particularly interesting to consider in light of standard "in-person" BDT formats versus newer computer-based formats that were used in some studies.)
- Practice effects within specific designs, i.e., does the test-taker take a different amount of time to place the first half of a given design vs. the second half?

Acknowledgments

This work was funded in part through the US NSF, awards #1936970, #2033413, and #2034013. We thank the anonymous reviewers for their detailed and helpful comments.

References

Akshoomoff, N. A., Delis, D. C., & Kiefner, M. G. (1989). Block constructions of chronic alcoholic and unilateral brain-damaged patients: A test of the right hemisphere vulnerability hypothesis of alcoholism. *Archives of Clinical Neuropsychology*, *4*(3), 275–281.

Begum, F. A., Begum, T., & Reza, F. (2017). Hand dominance and wais-r block design performance. *Journal of Advances in Medical and Pharmaceutical Sciences*, 1–5.

Ben-Yishay, Y., Diller, L., Mandleberg, I., Gordon, W., & Gerstman, L. J. (1971). Similarities and differences in block design performance between older normal and brain-injured persons. *J. Abnormal Psychology*, *78*(1), 17.

Burggraaf, R., Frens, M. A., Hooge, I. T., & van der Geest, J. N. (2016). A quick assessment of visuospatial abilities in adolescents using the design organization test (dot). *Applied Neuropsychology: Child*, *5*(1), 44–49.

Cha, S., Ainooson, J., Chong, E., Soulieres, I., Rehg, J. M., & Kunda, M. (2020). Enhancing cognitive assessment through multimodal sensing: A case study using the block design test. In *CogSci*.

Davis, R., Libon, D., Au, R., Pitman, D., & Penney, D. (2015). Think: Inferring cognitive status from subtle behaviors. *AI Magazine*, *36*(3), 49–60.

DeLuca, J., Kovaleski, M. E., Burright, R. G., & Donovick, P. J. (1990). Asymmetries in hand movement during block design construction. *Neuropsychologia*, *28*(7), 719–726.

Dirks, J. (1982). The effect of a commercial game on children's block design scores on the wisc-r iq test. *Intelligence*, *6*(2), 109–123.

Fenouillet, F., & Rozencwajg, P. (2015). Visual–spatial abilities and goal effect on strategies used to solve a block design task. *Learning & Individual Differences*, *39*, 158–163.

Frith, U., Happé, F., et al. (1995). Autism: Beyond "theory of mind.". *Cognition on cognition*, 13–30.

Hayhoe, M. M., Bensinger, D. G., & Ballard, D. H. (1998). Task constraints in visual working memory. *Vision research*, *38*(1), 125–137.

Hoffman, J. E., Landau, B., & Pagani, B. (2003). Spatial breakdown in spatial construction: Evidence from eye fixations in children with williams syndrome. *Cognitive Psychology*, *46*(3), 260–301.

Hutt, M. (1932). The kohs block-design tests. a revision for clinical practice. *J. Applied Psychology*, *16*(3), 298.

Jones, R. S., & Torgesen, J. K. (1981). Analysis of behaviors involved in performance of the block design subtest of the wisc-r. *Intelligence*, *5*(4), 321–328.

Joy, S., Fein, D., Kaplan, E., & Freedman, M. (2001). Quantifying qualitative features of block design performance among healthy older adults. *Archives of clinical neuropsychology*, *16*(2), 157–170.

Kohs, S. C. (1920). The block-design tests. *Journal of Experimental Psychology*, *3*(5), 357.

Kramer, J. H., Blusewicz, M. J., Kaplan, E., & Preston, K. A. (1991). Visual hierarchical analysis of block design configural errors. *Journal of Clinical and Experimental Neuropsychology*, *13*(4), 455–465.

Kunda, M. (2019). AI and cognitive testing: A new conceptual framework and roadmap. In *CogSci*.

Kunda, M., El Banani, M., & Rehg, J. M. (2016). A computational exploration of problem-solving strategies and gaze behaviors on the block design task. In *CogSci*.

Miller, J. C., Ruthig, J. C., Bradley, A. R., Wise, R. A., Pedersen, H. A., & Ellison, J. M. (2009). Learning effects in the block design task: A stimulus parameter-based approach. *Psychological assessment*, *21*(4), 570.

Paolo, A. M., & Ryan, J. J. (1994). Base rates of wais-r qualitative information for persons 75 years and older. *Assessment*, *1*(1), 83–88.

Pontius, A. (1993). Spatial representation, modified by ecology: From hunter-gatherers to city dwellers in indonesia. *Journal of cross-cultural psychology*, *24*(4), 399–413.

Pontius, A. (1995). In similarity judgments hunter-gatherers prefer shapes over spatial relations in contrast to literate groups. *Perceptual & motor skills*, *81*(3), 1027–1041.

Pontius, A. (1997). Spatial representation in face drawing and block design by nine groups from hunter-gatherers to literates. *Perceptual & motor skills*, *85*(3), 947–959.

Rozencwajg, P. (1991). Analysis of problem solving strategies on the kohs block design test. *European Journal of Psychology of Education*, *6*(1), 73.

Rozencwajg, P., Cherfi, M., Ferrandez, A., Lautrey, J., Lemoine, C., & Loarer, E. (2005). Age related differences in the strategies used by middle aged adults to solve a block design task. *The International Journal of Aging and Human Development*, *60*(2), 159–182.

Rozencwajg, P., & Corroyer, D. (2001). Strategy development in a block design task. *Intelligence*, *30*(1), 1–25.

Rozencwajg, P., & Fenouillet, F. (2012). Effect of goal setting on the strategies used to solve a block design task. *Learning & Individual Differences*, *22*(4), 530–536.

Salthouse, T. (1987). Sources of age-related individual differences in block design tests. *Intelligence*, *11*(3), 245–262.

Schatz, A. M., Ballantyne, A. O., & Trauner, D. A. (2000). A hierarchical analysis of block design errors in children with early focal brain damage. *Developmental Neuropsychology*, *17*(1), 75–83.

Schorr, D., Bower, G. H., & Kiernan, R. J. (1982). Stimulus variables in the block design task. *Journal of consulting and clinical psychology*, *50*(4), 479.

Stewart, M. E., Watson, J., Allcock, A.-J., & Yaqoob, T. (2009). Autistic traits predict performance on the block design. *Autism*, *13*(2), 133–142.

Troyer, A. K., Cullum, C. M., Smernoff, E. N., & Kozora, E. (1994). Age effects on block design: Qualitative performance features and extended-time effects. *Neuropsychology*, *8*(1), 95.

Vasterling, J. J., Rogers, C., & Kaplan, E. (2000). Qualitative block design analysis in posttraumatic stress disorder. *Assessment*, *7*(3), 217–226.