



DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

The Design Organization Test: Further Demonstration of Reliability and Validity as a Brief Measure of Visuospatial Ability

The Harvard community has made this article openly available.
[Please share](#) how this access benefits you. Your story matters.

Citation	Killgore, William D. S., and Hannah Gogel. 2014. "The Design Organization Test: Further Demonstration of Reliability and Validity as a Brief Measure of Visuospatial Ability." <i>Applied Neuropsychology</i> . Adult 21 (4): 297-309. doi:10.1080/23279095.2013.811671. http://dx.doi.org/10.1080/23279095.2013.811671 .
Published Version	doi:10.1080/23279095.2013.811671
Accessed	February 17, 2015 6:30:22 AM EST
Citable Link	http://nrs.harvard.edu/urn-3:HUL.InstRepos:13454784
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

(Article begins on next page)

The Design Organization Test: Further Demonstration of Reliability and Validity as a Brief Measure of Visuospatial Ability

William D. S. Killgore

Center for Depression, Anxiety, and Stress Research, McLean Hospital, Belmont, Massachusetts and Department of Psychiatry, Harvard Medical School, Boston, Massachusetts

Hannah Gogel

Center for Depression, Anxiety, and Stress Research, McLean Hospital, Belmont, Massachusetts

Neuropsychological assessments are frequently time-consuming and fatiguing for patients. Brief screening evaluations may reduce test duration and allow more efficient use of time by permitting greater attention toward neuropsychological domains showing probable deficits. The Design Organization Test (DOT) was initially developed as a 2-min paper-and-pencil alternative for the Block Design (BD) subtest of the Wechsler scales. Although initially validated for clinical neurologic patients, we sought to further establish the reliability and validity of this test in a healthy, more diverse population. Two alternate versions of the DOT and the Wechsler Abbreviated Scale of Intelligence (WASI) were administered to 61 healthy adult participants. The DOT showed high alternate forms reliability ($r = .90-.92$), and the two versions yielded equivalent levels of performance. The DOT was highly correlated with BD ($r = .76-.79$) and was significantly correlated with all subscales of the WASI. The DOT proved useful when used in lieu of BD in the calculation of WASI IQ scores. Findings support the reliability and validity of the DOT as a measure of visuospatial ability and suggest its potential worth as an efficient estimate of intellectual functioning in situations where lengthier tests may be inappropriate or unfeasible.

Key words: IQ, neuropsychology, reliability, validity, visuospatial ability

INTRODUCTION

Neuropsychological assessment serves a vitally important function within a broad range of research and clinical settings. Whereas neuroimaging can provide important data regarding brain abnormalities and functional localization, only the neuropsychological assessment can provide standardized, reliable, and valid information about the individual's functional capabilities. Because of the complexity of human cognitive functioning, traditional neuropsychological assessments have often required long testing sessions, sometimes lasting many hours

© William D. S. Killgore and Hannah Gogel

This is an Open Access article. Non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly attributed, cited, and is not altered, transformed, or built upon in any way, is permitted. The moral rights of the named authors have been asserted.

Address correspondence to William D. S. Killgore, Ph.D., Center for Depression, Anxiety, and Stress Research, McLean Hospital, Harvard Medical School, 115 Mill Street, Belmont, MA 02478. E-mail: Killgore@mclean.harvard.edu

(Lezak, Howieson, & Loring, 2004). In many cases, it is not unheard of for a comprehensive neuropsychological assessment battery to take the better part of a full day for completion. In recent years, considerable effort has been devoted toward reducing the administration time of many neuropsychological tests (Donders, 2001; Farias et al., 2011; Meulen et al., 2004; Sunderland, Slade, & Andrews, 2012). Long sessions of testing can cause examinee fatigue, which can be particularly trying for patients already compromised by neurological or medical conditions (Lezak et al., 2004). Not only are lengthy testing sessions unpleasant for examinees, but they can also reduce the reliability and validity of the obtained data. Furthermore, longer assessments are generally more expensive as well, a problem that has further fueled efforts to develop more time-efficient methods for neuropsychological assessment (Groth-Marnat, 2000; Ricker, 1998; Sunderland et al., 2012). Overall, the emerging consensus in the field is that assessments that obtain the same information in less time are to be preferred (Donders, 2001).

Standardized measures of intellectual functioning, such as the various Wechsler intelligence scales, often take 1 hr to 2 hr to administer in some clinical settings. Abbreviated versions, such as the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999), although considerably less extensive, can still take 30 min to 1 hr to administer in actual clinical practice. Even isolated subscales, such as the Block Design (BD) subtest, can take a quarter of an hour or more to administer to some slower examinees. In an effort to provide a rapid screening measure for visuospatial ability comparable to the BD subtest of the Wechsler scales, we developed a brief paper-and-pencil measure, the Design Organization Test (DOT; Killgore, Glahn, & Casasanto, 2005). The DOT is a single-page manually administered test that presents visual stimuli similar to those used by the Wechsler BD subtests but utilizes a briefer paper-and-pencil completion format. The DOT is relatively quick, requiring only 2 min to administer, and can save considerable time compared with the complete administration of the BD subtest by eliminating the need to exhaust time limits on multiple-item administrations. From the initial validation studies (Killgore et al., 2005), the DOT was shown to have good alternate forms reliability ($r = .80$) in a large sample of college students and was found to correlate very highly with the actual BD subtest of the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; $r = .92$) in a sample of neurologic clinic patients undergoing neuropsychological assessment. Moreover, in that study, when DOT scores were converted to BD scores using regression procedures and were replaced in the calculation of Full-Scale IQ (FSIQ) and Performance IQ (PIQ) scores, it was found that scores only differed from the traditional method by about half of a

single IQ point on average. Thus, the preliminary validation studies suggest that the DOT may provide comparable information to the WAIS-III BD subtest, but in significantly less time and with fewer cumbersome test materials.

Although the initial validation studies are encouraging (Killgore et al., 2005), further validation of the DOT is necessary to establish its clinical utility. For instance, the first study in the series examining the reliability of the DOT was performed with a sample of 411 young-adult students (average age = 18.0 years, $SD = 1.0$) attending a highly selective private university, and the findings may therefore overestimate performance on the DOT. The second sample included only neurological patients with documented brain lesions or neurologic impairment (average age = 47.7 years, $SD = 15.3$), potentially underestimating DOT performance in the general population. Consequently, the present study provides additional reliability and validity data for the DOT obtained from a sample of healthy nonclinical participants with a broader range of intellectual functioning than in the prior reports.

METHOD

Participants

Sixty-one healthy right-handed adults (30 men; 31 women), recruited from flyers and Internet advertisements within the greater Boston area, participated in this study. All participants spoke English as their primary language and were aged 18 to 45 years old ($M_{\text{age}} = 30.3$ years, $SD = 8.1$). Participants were initially screened to exclude any history of psychopathology, serious medical conditions, neurologic conditions, alcohol or substance abuse/dependence, use of illicit substances, or head injury. The final sample was racially diverse and included 69% Caucasian, 15% African American, 10% Asian, 3% Other race, and 3% Multiracial background. Additionally, 5% of participants also coendorsed Hispanic/Latino heritage. Years of formal education ranged from 11 to 20 ($M = 14.9$ years, $SD = 2.2$). Regarding the highest education level attained, 3.3% had less than a high school diploma, 18% percent of participants terminated education with a high school diploma, 31.2% had completed some college, 23% had attained an undergraduate degree, 13.1% had some postgraduate education without a degree, 9.8% had obtained a master's degree, and 1.6% had completed doctoral studies. All participants provided written informed consent before enrollment in the study. This research protocol was approved by the McLean Hospital Institutional Review Board and the U.S. Army Human Research Protection Office.

Materials

Wechsler Abbreviated Scale of Intelligence (Pearson Assessment, Inc., San Antonio, TX). The WASI (Wechsler, 1999) was administered to evaluate the validity of the DOT as a measure of visuospatial ability and as a potential brief surrogate for the BD subtest when determining IQ scores. The WASI is a commonly used brief-format intelligence scale, which has been found to have a .92 correlation with the more extensive WAIS-III (Pearson Assessment, Inc., San Antonio, TX) and is reported to have a .98 reliability for FSIQ. A post-baccalaureate research technician, supervised by a doctoral-level neuropsychologist, administered all tests.

Design Organization Test (Killgore et al., 2005). Participants completed two alternate versions of the DOT (Killgore et al., 2005) in counterbalanced order. The two alternate forms of the DOT are provided in the Appendix, along with the standard practice page and scoring key. At the outset of each administration, participants completed the practice page, which displayed a code key comprising six squares, each with unique black and white patterns of shading and a corresponding identification number placed immediately above. Below the code key was a fully completed demonstration example and two incomplete practice items to be filled in by the participant.

Two alternate forms of the DOT were used (Form A and Form B). Each version included nine square designs (five small designs and four large designs). Below each design was an identically sized empty square grid. The small designs were paired with 2×2 grids, whereas the large designs were paired with 3×3 grids. Each grid square corresponded to a location on the associated design above it and served as a response blank that could be completed by entering the corresponding number from the code key located at the top of the page.

Procedures

Participants were administered both versions (Form A and Form B) of the DOT during a larger test battery including various paper-and-pencil and computerized cognitive tasks. The two DOT versions were administered approximately 30 min apart from one another, separated by other cognitive tasks, and no sooner than 15 min after the full WASI administration. The DOT was always administered after the WASI, so as not to bias WASI responses via potential problem-solving strategies that might have been discovered through exposure to the DOT.

The order of the DOT administration was counterbalanced, with 31 participants (51%) completing Form A first (i.e., order AB) and 30 participants (49%) completing Form B first (i.e., order BA). To begin the test, the administrator placed the practice page directly in front of the participant (see the Appendix), pointed to the key at the top of the sheet, and read the following instructions aloud to the participant:

Look at these six boxes. Each box is different. Every box has a different design and has its own number from 1 to 6. Number 1 is solid black. Number 2 is half black and half white. So is Number 3, but if you look closely, you'll see that the design for Box Number 3 is different from Number 2. All six boxes are different, and each one has its own number.

The administrator then pointed to the example items below the key and said:

Now look down here. This square design is made up of four of the boxes I just showed you. Down below are empty squares for you to put the numbers that match each box. This first one is already done for you. See, the first square is solid black. Look at the code key; the number for all solid black squares is Number 1. So number "1" goes in this square. The lower right box is also black, so I also put a "1" in that square too. Here are two white boxes. See, the solid white boxes are always Number 6, so the number "6" goes here and here.

Pointing to the next sample, the administrator said:

Over here is a set of empty squares. See if you can match the design by filling the code numbers. See, the first box is solid white. So what number would you put down here to match? Go ahead and complete the rest of the practice items.

After the participant had completed the practice section in pencil, the administrator said:

So all you have to do is put the correct number in each of the empty boxes to match the pattern above it. On the next page, there are many designs much like the ones we just did. Look at each design and fill in as many of the empty boxes below it as you can with the correct numbers to match the designs. Do as many as you can. You will have only 2 min, so work as quickly as you can. Do you have any questions?

The administrator turned over the page and started the timer. The participant was allowed to complete as many of the DOT items as possible in 2 min. Although not explicitly stated in the instructions, participants were not required to complete the items in any particular order and were permitted to use their pencil to make marks or lines on the target items as they wished. Exactly at 2 min, the participant was instructed to stop working and put down the pencil.

If during the administration, a participant asked if the items could be completed out of order or if it was permissible to draw on the items, the administrator would reply, "Yes, you may complete the task any way you wish." Each administration of the DOT started with complete instructions and a practice section.

RESULTS

WASI Performance

Performance on the WASI covered a broad range of ability levels. The WASI FSIQ of participants ranged from 71 to 138 ($M = 111.1, SD = 16.2$). Verbal IQ (VIQ) ranged from 64 to 133 ($M = 110.4, SD = 15.8$), while PIQ ranged from 67 to 134 ($M = 108.8, SD = 15.6$).

Normative Data and Equivalence of Alternate Forms of the DOT

The two versions of the DOT (i.e., Form A and Form B) were administered in a counterbalanced order, with essentially half of the sample receiving Form A first and the other half receiving Form B first. Table 1 presents the mean scores for each version, administered either as the first or second in the series. When administered first in the series, Form A and Form B produced nearly identical mean scores, $t(59) = 0.20, p = .84$ (see Table 1). Similarly, when administered second in the series, the mean scores for Form A and Form B did not differ significantly, $t(59) = 0.63, p = .53$, suggesting comparability in the scores provided by the two forms. There were no significant sex differences in performance (see Table 1), with men and women performing similarly at the first, $t(59) = 0.91, p = .37$, and second administrations, $t(59) = 0.92, p = .36$.

Reliability of Alternate Forms of the DOT

The two versions of the DOT were also highly correlated (see Figure 1). When Form A was followed by Form B,

TABLE 1
Mean Scores for Form A and B of the DOT at Each Administration

Test	Administered First <i>M (SD)</i>	Administered Second <i>M (SD)</i>
DOT A then B	(Form A) 35.90 (8.06)	(Form B) 41.03 (9.10)
DOT B then A	(Form B) 35.43 (10.05)	(Form A) 39.43 (10.61)
Overall Mean	35.67 (9.02)	40.25 (9.82)
Men	34.60 (8.86)	39.07 (9.80)
Women	36.71 (9.19)	41.39 (9.86)

Note. Forms were administered in counterbalanced order with 30 min between administrations. Approximately half of the sample ($n = 31$) completed Form A followed by Form B, while the remainder ($n = 30$) completed Form B followed by Form A.

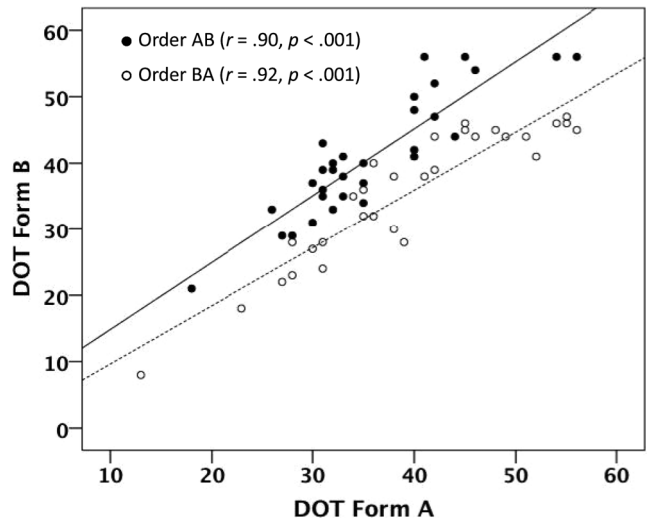


FIGURE 1 Scatterplot showing the relationship between alternate forms of the DOT. The black circles (solid line) show the association between forms when Form A was completed before Form B (i.e., order AB). The empty white circles (dashed line) show the association when Form B was completed before Form A (i.e., order BA). The difference between the parallel regression lines reflects the effects of practice between the two administrations.

the correlation between the two versions was significant. When Form B was followed by Form A, the correlation was similarly high. Finally, when performance on the first administration was correlated with performance on the second, regardless of DOT version, scores on the two administrations remained highly associated, $r(59) = .91, p < .001$, suggesting excellent alternate forms and test-retest reliability.

Practice Effects

There was evidence of modest but significant improvement between the two administrations of the alternate forms of the DOT. As shown in Table 1, participants scored an average of 35.67 points ($SD = 9.02$) at the first administration (without regard to version), but this increased to 40.25 points ($SD = 9.82$) by the second administration, which was completed 30 min later. On average, participants showed a within-subject improvement of 4.57 points between administrations across the two forms, $t(60) = 8.81, p < .001$. Practice effects were similar regardless of whether Form A preceded Form B ($M = 5.13, SD = 4.05$) or if Form B preceded Form A ($M = 4.00, SD = 4.05$), $t(59) = 1.09, p = .28$.

Scoring Issues

Commission errors. Consistent with the published instructions (Killgore et al., 2005), the total score for the DOT was calculated by simply counting the number

TABLE 2
Frequency of Commission Errors on the DOT at Each Administration

Number of Errors	First Administration			Second Administration		
	Freq.	%	%ile	Freq.	%	%ile
0	36	59.0	59.0	38	62.3	62.3
1	12	19.7	78.7	9	14.8	77.0
2	5	8.2	86.9	3	4.9	82.0
3	3	4.9	91.8	6	9.8	91.8
4	1	1.6	93.4	—	—	—
5	—	—	—	1	1.6	93.4
6	—	—	—	—	—	—
7	—	—	—	2	3.3	96.7
8	2	3.3	96.7	—	—	—
9	—	—	—	2	3.3	100.0
10	1	1.6	98.4	—	—	—
11	1	1.6	100.0	—	—	—

Note. Freq. = Frequency; % = percent of sample with the specified number of errors; %ile = cumulative percentage of the sample scoring at or below the specified number of errors.

of correctly completed response squares, regardless of location on the page or order of completion. There was no penalty for incorrect responses (i.e., entering the wrong response in a box). On the whole, participants made very few errors of this type on either the first ($M=1.18$, $SD=2.38$, range=0–11) or second ($M=1.15$, $SD=2.16$, range=0–9) administrations of the DOT, and the number of errors did not differ significantly, $t(60)=0.15$, $p=.88$, between administrations. As evident in Table 2, the majority of participants made no commission errors at all, and it was rare to find more than three errors in either administration of the DOT.

Use of strategy. In prior experience with the DOT, it was discovered that some respondents self-implement a strategy of drawing grid squares on the target designs to help delineate the boundaries of the individual design elements. Because the instructions to the participant do not explicitly forbid this strategy, it was of interest to determine whether it influenced performance. On the first administration, 10 participants (16.4%) employed this strategy without prompting. Overall, participants using the strategy on the first administration ($M=38.80$, $SD=5.90$) did not score significantly differently compared with those not using the strategy ($M=35.06$, $SD=9.43$) $t(59)=1.20$, $p=.23$. By the second administration of the DOT, 14 participants (23.0%) spontaneously chose to draw grids on some of the designs (9 of the 10 participants choosing this strategy on the first administration chose to use this strategy again on the second administration). Again, the performance of those using the strategy ($M=41.21$, $SD=9.78$) was not significantly different from those not employing the strategy ($M=39.96$, $SD=9.92$) $t(59)=0.42$, $p=.68$. Hence, we conclude that use of the

grid-drawing strategy appears to have no appreciable effect on performance, but given the modest power in the present sample, this would be an important topic for further research.

Ceiling effect. Because the DOT contains only 56 items to be completed within 120 s, it is possible for some participants to complete all items within the allotted time, leading to an upper limit (i.e., ceiling effect) for performance. This is evident by a slight negatively skewed distribution on the first (skew = $-.37$, $SD=0.31$) and second (skew = $-.14$, $SD=0.31$) administrations of the DOT. On the first administration of the DOT, one participant (1.6%) correctly completed all 56 items of the DOT within the 2-min time period. By the second administration, five participants (8.2%) correctly completed all 56 items of the DOT within the allotted time. Thus, even when two versions of the DOT are completed within close temporal proximity,

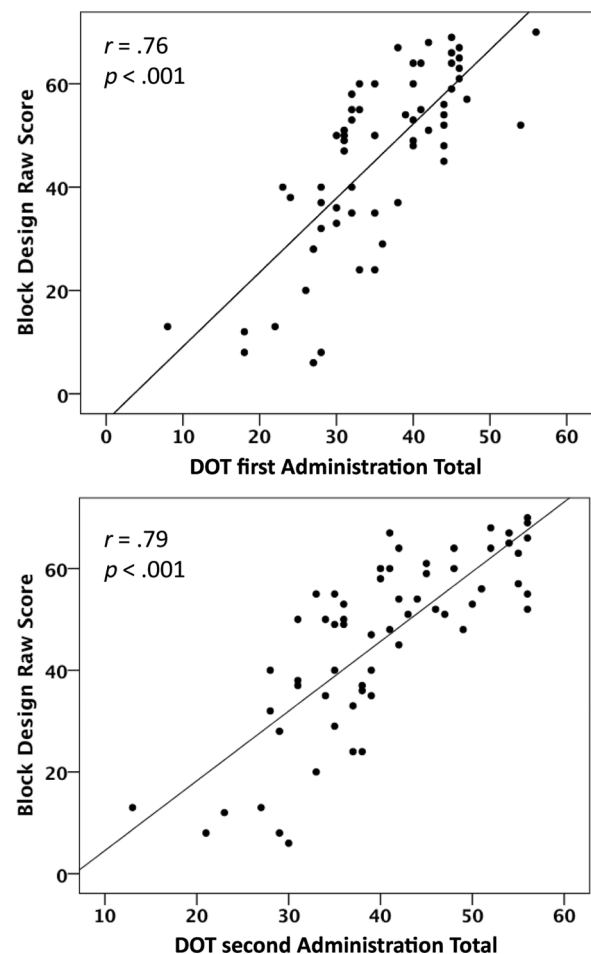


FIGURE 2 Scatterplots showing the association between the first (top panel) and second (bottom panel) DOT administrations and raw WASI Block Design scores.

TABLE 3
Intercorrelations Between DOT and Criterion Measures

Criterion Test	DOT Time 1	DOT Time 2	Raw Block Design
Age	-.17	-.23	-.03
Education	.30*	.33*	.36**
WASI 4-Test FSIQ	.68***	.69***	.86***
WASI 2-Test FSIQ	.61***	.62***	.74***
WASI VIQ	.57***	.57***	.67***
WASI PIQ	.68***	.69***	.92***
WASI Vocabulary	.58***	.59***	.63***
WASI Similarities	.49***	.48***	.61***
WASI Block Design	.73***	.75***	.99***
WASI Matrix Reasoning	.50***	.49***	.69***

Note. All WASI scores are age-corrected standard scores.
* $p < .05$. ** $p < .005$. *** $p < .001$.

only truly exceptional performers are likely to achieve the maximum score.

Concurrent Validity

To establish the concurrent validity of the DOT, performance on the first and second administrations was correlated with demographic variables of age and education

and the various age-corrected standardized scales of the WASI. Because a primary goal of the study was to show the usefulness of the DOT as a surrogate paper-and-pencil measure for BD, we included the raw BD scores from the WASI as a comparator variable. As evident in Figure 2, scores on the first and second administrations of the DOT were highly correlated with raw BD scores. Table 3 shows that DOT scores were significantly correlated with all of the subtests of the WASI, as well as the FSIQ. In contrast, similar to raw BD scores, the DOT was not correlated with age and was only weakly correlated with education.

Construct Validity

Construct validity is bolstered by the differences in mean scores obtained in the various samples tested across studies. The present sample mean for the first administration (see Table 1) is significantly lower, $t(470) = 7.62$, $p < .0001$, than that obtained in our prior study of young students at a highly selective private university ($M = 44.00$, $SD = 7.80$, $n = 411$; Killgore et al., 2005) and is significantly higher, $t(100) = 5.27$, $p < .0001$, than that obtained by a clinical sample of neurological patients with documented brain lesions ($M = 24.32$,

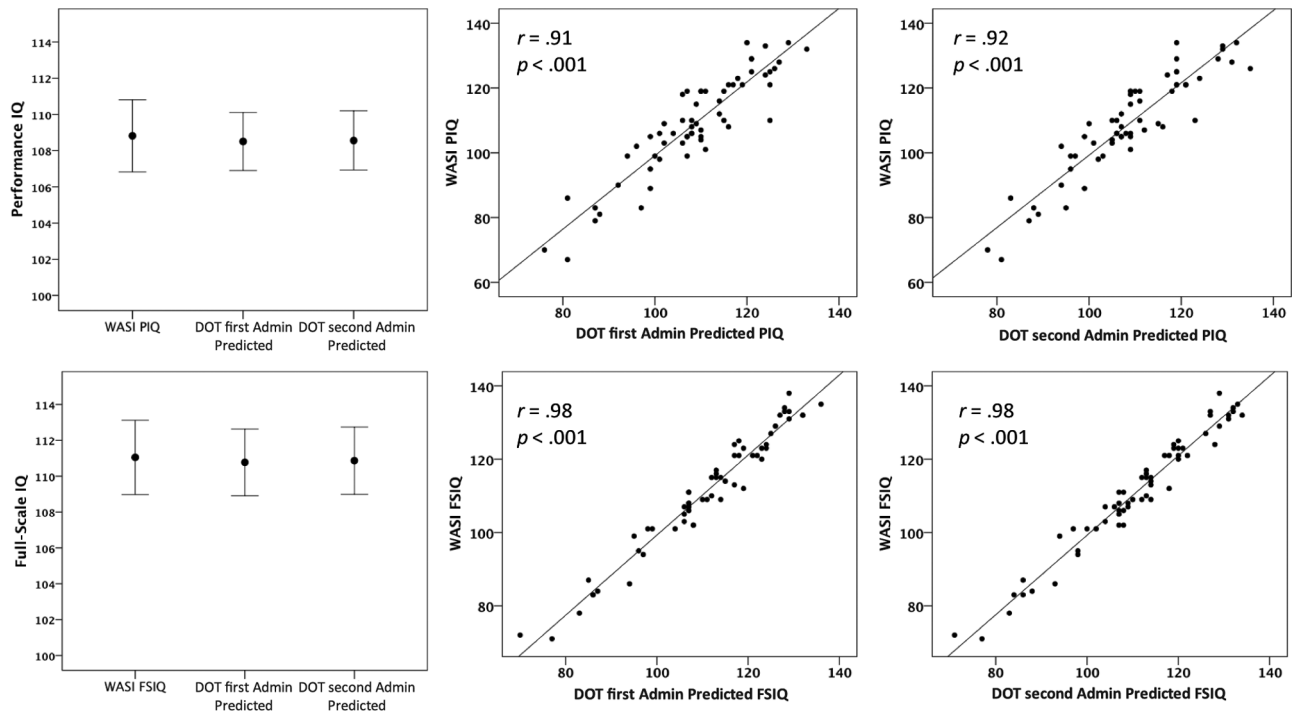


FIGURE 3 Effects of replacing WASI Block Design (BD) scores with predicted BD scores derived from the DOT. Top Row: Performance IQ (PIQ) scores (Mean \pm 1 SE) derived from either the first or second administration of the DOT did not differ significantly from actual WASI PIQ scores (left panel). Actual WASI PIQ scores were highly correlated with PIQ scores derived from the first (middle panel) and second (right panel) administrations of the DOT. Bottom Row: Full-Scale IQ (FSIQ) scores (Mean \pm 1 SE) derived from either the first or second administration of the DOT did not differ significantly from actual WASI FSIQ scores (left panel). Actual WASI FSIQ scores were highly correlated with FSIQ scores derived from the first (middle panel) and second (right panel) administrations of the DOT.

$SD=12.75$, $n=41$) described in the prior report (Killgore et al., 2005). These differences suggest that the DOT is sensitive to group differences in ability level, ranging from impaired patients, to healthy participants, to exceptionally bright students at a top competitive university.

Usefulness as a Surrogate for BD

One goal of the present study was to examine the usefulness of the DOT as a surrogate for the BD subtest in calculating WASI IQ scores. To test this, we first estimated BD subtest scores through a series of 61 linear regression analyses using a sequential “leave-one-out” jackknife resampling method with DOT performance as the predictor. For each regression analysis, a different participant was excluded from the sample and the DOT scores from the remaining 60 participants were used to predict the raw BD score of the excluded participant. All of these equations yielded R values of .74 to .80 and all were significant ($ps < .001$). The estimated raw BD scores based on these equations were then used in place of actual BD scores to recalculate derived PIQ and FSIQ scores for each participant. This process was repeated for the first and second administrations of the DOT.

As shown in Figure 3, actual mean PIQ scores ($M=108.8$, $SD=15.6$) did not differ significantly from the estimated PIQ scores derived from either the first ($M=108.5$, $SD=12.5$), $t(60)=0.37$, $p=.71$, or second ($M=108.6$, $SD=12.8$), $t(60)=0.32$, $p=.75$, administrations of the DOT. As shown in Figure 3, PIQ scores estimated from the DOT were highly correlated with actual PIQ scores. Similarly, actual FSIQ scores ($M=111.1$, $SD=16.2$) did not differ significantly from those estimated from the first ($M=110.8$, $SD=14.5$), $t(60)=-0.59$, $p=.56$, and second ($M=110.9$, $SD=14.6$), $t(60)=0.41$, $p=.69$, administrations of the DOT. Again, FSIQ scores estimated from the DOT were highly correlated with actual FSIQ scores (see Figure 3).

DISCUSSION

The goal of the present study was to provide additional data regarding the reliability and validity of the DOT within a healthy nonclinical sample. Presently, the two alternate forms of the DOT yielded nearly identical mean scores, showed similar practice effects, and demonstrated extremely high alternate forms reliability. DOT performance was not significantly affected by the presence or absence of errors of commission, use of strategic marking on the stimuli, or ceiling effects, although further study of these issues is warranted. The present study also provided further confirmation of the concurrent validity of the DOT as a measure of visuospatial

ability and its usefulness as a potential surrogate for the BD subtest in estimating IQ scores when assessment time is limited. Because the DOT can be administered in only 2 min and without additional materials, these findings suggest that it may serve as a brief screening instrument for rapidly evaluating visuospatial ability or estimating IQ under conditions where a lengthier assessment might not be feasible or appropriate.

For speed tests such as the DOT, measures of internal consistency reliability are not considered appropriate (Nunnally, 1978). Instead, for timed tests that emphasize speed, test-retest methods or the evaluation of performance on alternate forms can provide an indication of the reliability of the instrument. We found that the two alternate forms of the DOT were highly correlated, suggesting excellent reliability. Overall, the reliability coefficient obtained presently ($r=.91$) was higher than previously reported in the initial validation study in a university student sample ($r=.80$; Killgore et al., 2005). This difference is likely due to the fact that the college student participants in the initial study were run in large groups of several hundred at a time, whereas the data in the present study were collected one on one with the test administrator, which may have improved participant compliance with all of the study procedures and served to enhance overall reliability. The current reliability coefficients are also similar to the test-retest reliability reported by the test publisher for the WASI BD ($r=.92$; Wechsler, 1999), and they actually exceed the reliability reported for the WAIS-III BD subtest ($r=.82$; Wechsler, 1997). From these findings, we conclude that the DOT provides at least comparable reliability to the considerably more time-consuming and labor-intensive BD subtest used in the Wechsler scales.

Consistent with the previously published validation studies (Killgore et al., 2005), the present findings suggest that the two alternate forms of the DOT are indeed interchangeable and provide comparable scores. Consequently, the mean score we report across the two forms can be confidently used as a normative comparator for either version. For most uses among adults aged 18 to 45 years old, we recommend the normative data obtained from the present sample ($M=35.67$, $SD=9.02$), as it was obtained from a broad range of healthy individuals across the normal bounds of intellectual capacity. However, we also compared the present mean performance to that reported in the prior validation studies. The mean score on the DOT from the present sample falls midrange between the mean obtained for young students at a highly selective private university and scores obtained from a clinical sample of patients with documented neurological disorders (Killgore et al., 2005). The fact that DOT scores differ across these very different populations provides further support for the

construct validity of the test as a measure of visuospatial and intellectual ability.

Concurrent validity of the DOT was demonstrated by high intercorrelations with WASI BD and all WASI IQ scales. Age was not significantly correlated with DOT performance in the present study, a finding that differed from the initially reported validation in the sample of patients from a neurology clinic. In the prior study (Killgore et al., 2005), age was significantly negatively correlated with DOT and BD performance. However, this difference is likely due to restriction of range, as the present study was significantly younger and ranged from 18 to 45 years old, whereas the neurologic sample studied in our prior report ranged from 18 to 76 years of age. It is likely that age effects on DOT scores are minimal among young to middle-age individuals but may have a greater influence among middle-age to elderly participants and those with some form of neuropathology. This is an important consideration in actual clinical settings, as the current normative data would be inappropriate for application with individuals older than 45 years old because it would likely lead to incorrect interpretations. Further research with the DOT in older populations is encouraged to establish valid normative ranges and to document the potential changes in performance with advancing age.

To facilitate the potential use of the DOT as a surrogate measure for BD during WASI administration, we employed a “leave-one-out” jackknife resampling regression analysis to permit unbiased prediction of raw BD scores from the DOT. WASI IQ scores were then recalculated using these predicted BD scores and compared to the original IQ scores calculated using actual raw BD scores. Replacing the raw BD scores with the estimated scores derived from the DOT yielded virtually no difference between estimated and actual IQ scores (i.e., actual and estimated scores differed by about one third of a single IQ point). This replicates the previous finding from the initial validation study, which found that using an estimated BD score derived from the DOT in place of actual BD scores had virtually no effect on PIQ or FSIQ from the WAIS-III (Killgore et al., 2005). As reported in the previous study, estimated and actual IQ scores differed by less than half of a single point on average. These findings suggest that with additional validation, the DOT may prove useful as a rapidly administered surrogate for BD under some conditions. We encourage additional research into this issue with larger samples and a broader age range.

The present data support the potential utility of the DOT as a brief and rapidly administered instrument for assessing visuospatial ability and estimating general intellectual ability. Since 2009, the DOT has been included as part of the large-scale Rotterdam Study in the Netherlands (Hofman et al., 2011), which

longitudinally tracks a number of health-related issues including neurological status among a cohort of 14,926 participants older than the age of 45. As part of the Rotterdam Study, a recent report examined DOT scores as one outcome measure in a study of the effects of chemotherapy on cognitive functioning in elderly Dutch women with breast cancer (Koppelmans et al., 2012). The women in that study ranged in age from 50 to 80 years old and had undergone adjuvant chemotherapy at least 20 years earlier. Chemotherapy had no effect on DOT performance or several other timed neuropsychological tasks such as verbal fluency, the Purdue Pegboard Test, and Stroop Word Naming, but it did affect measures of verbal memory and Stroop Interference. Interestingly, women in both the chemotherapy group ($M = 28.9$, $SD = 9.2$, $n = 195$) and the reference group ($M = 28.9$, $SD = 9.7$, $n = 511$) had DOT scores that were significantly lower than those observed in our current sample. The lower scores are likely to be attributable to the significantly older age among the women in that sample compared with the present group.

The reliability and validity of the DOT appears comparable to other relatively brief tests purported to measure visuospatial ability and estimate intellectual functioning. For instance, an early study (Hall, 1957) compared the 30-item Raven's Progressive Matrices, a nonverbal test of intelligence, with scores on the original WAIS and found that scores correlated with PIQ ($r = .71$), VIQ ($r = .58$), and FSIQ ($r = .72$). Others have reported correlations with other standard intelligence tests near .70 (Burke, 1985; Jensen, Saccuzzo, & Larsen, 1988; O'Leary, Rusch, & Guastello, 1991), with test-retest reliability generally greater than .80 (Burke, 1985). Overall, the current findings for the DOT are quite similar to those reported for the other instruments (see Table 3).

There are a couple of important test administration issues that emerge from this study. The first is the potential role of ceiling effects on the DOT. We found that a very small percentage of our sample (one participant; 1.6% of the sample) was able to complete the entire task within the 2-min time limit, and this percentage was slightly larger (five participants; 8% of the sample) following a second administration with an alternate version. The one individual obtaining this performance on the first administration had a measured FSIQ of 135, while the mean FSIQ of the five participants achieving ceiling on the second administration was 124, suggesting that this ceiling is likely to be achieved only by exceptionally capable individuals. However, these findings suggest that although the DOT is likely to be appropriate for clinical populations and for general assessment of visuospatial deficits, it would be inappropriate for testing the higher limits of visuospatial ability and intellectual capacity due to this potential ceiling effect. The

second issue involves the fact that a minority of participants spontaneously drew lines on the designs to demarcate boundaries. We intentionally did not restrict participants from drawing on the form to evaluate the potential effect of this strategy. Our results showed very minimal and nonsignificantly higher scores among those employing this strategy. However, the present sample was modest in size, so it is not possible to know whether such a strategy might have had an effect in a larger sample. Furthermore, it is not clear whether better scores would result from the strategy itself, or if they would merely reflect that more capable individuals were more likely to spontaneously think of employing such a strategy. To definitively answer this question, it will be necessary for future research to randomly assign large numbers of participants to groups required to use or not use the line-drawing strategy.

The present study adds important information regarding the reliability and validity of the DOT in a sample of healthy volunteers. However, the present sample is limited in size, and further normative studies with a variety of populations, larger samples, and broader age range are necessary to establish the DOT as widely useful clinical tool. Research on the DOT has been limited because the actual test forms were never published in our prior study and were, therefore, not readily available to many clinicians and researchers. To facilitate further research and clinical use with the DOT, both alternate forms (A and B) are provided as an Appendix to this article. Preliminary findings from this and prior studies suggest that the DOT provides a reliable and valid assessment of visuospatial ability that can be obtained in a little more than 2 min, with no need for cumbersome equipment or test materials. Furthermore, with further research and validation, the DOT may be useful for estimating general intellectual ability when time constraints or other circumstances hinder the ability to obtain a more comprehensive assessment.

ACKNOWLEDGEMENTS

This research was supported by a United States Army Medical Research Acquisition Activity grant (W81XWH-09-1-0730). The authors wish to thank Zachary J. Schwab, Melissa R. Weiner, Sophie R. DelDonno, and Maia Kipman for their assistance in data collection.

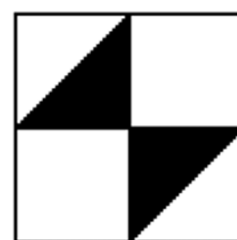
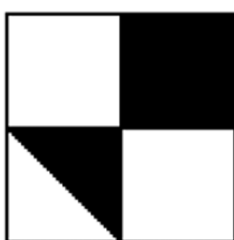
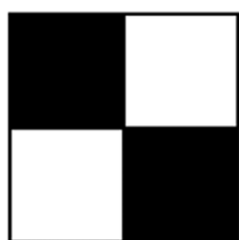
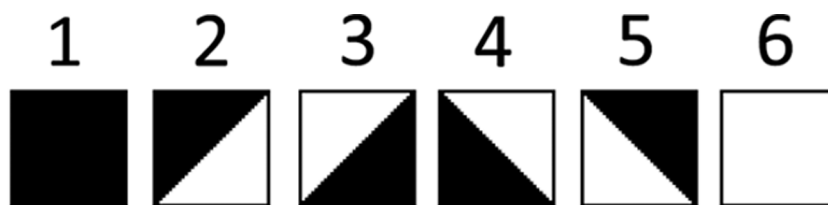
REFERENCES

- Burke, H. R. (1985). Raven's Progressive Matrices (1938): More on norms, reliability, and validity. *Journal of Clinical Psychology, 41*, 231–235.
- Donders, J. (2001). Using a short form of the WISC-III: Sinful or smart? *Child Neuropsychology, 7*, 99–103.
- Farias, S. T., Mugas, D., Harvey, D. J., Simmons, A., Reed, B. R., & DeCarli, C. (2011). The measurement of everyday cognition: Development and validation of a short form of the Everyday Cognition scales. *Alzheimer's & Dementia, 7*, 593–601.
- Groth-Marnat, G. (2000). Visions of clinical assessment: Then, now, and a brief history of the future. *Journal of Clinical Psychology, 56*, 349–365.
- Hall, J. (1957). Correlation of a modified form of Raven's Progressive Matrices (1938) with the Wechsler Adult Intelligence Scale. *Journal of Consulting Psychology, 21*, 23–26.
- Hofman, A., van Duijn, C. M., Franco, O. H., Ikram, M. A., Janssen, H. L., Klaver, C. C., ... Wittman, J. C. (2011). The Rotterdam Study: 2012 objectives and design update. *European Journal of Epidemiology, 26*, 657–686.
- Jensen, A. R., Saccuzzo, D. P., & Larsen, G. E. (1988). Equating the standard and advanced form of the Raven Progressive Matrices. *Educational and Psychological Measurement, 48*, 1091–1095.
- Killgore, W. D., Glahn, D. C., & Casasanto, D. J. (2005). Development and validation of the Design Organization Test (DOT): A rapid screening instrument for assessing visuospatial ability. *Journal of Clinical and Experimental Neuropsychology, 27*, 449–459.
- Koppelmans, V., Breteler, M. M., Boogerd, W., Seynaeve, C., Gundy, C., & Schagen, S. B. (2012). Neuropsychological performance in survivors of breast cancer more than 20 years after adjuvant chemotherapy. *Journal of Clinical Oncology, 30*, 1080–1086.
- Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment*. New York, NY: Oxford University Press.
- Meulen, E. F., Schmand, B., van Campen, J. P., de Koning, S. J., Ponds, R. W., Scheltens, P., & Verhey, F. R. (2004). The seven minute screen: A neurocognitive screening test highly sensitive to various types of dementia. *Journal of Neurology, Neurosurgery & Psychiatry, 75*, 700–705.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.
- O'Leary, U.-M., Rusch, K. M., & Guastello, S. J. (1991). Estimating age-stratified WAIS-R IQs from scores on the Raven's Standard Progressive Matrices. *Journal of Clinical Psychology, 47*, 277–284.
- Ricker, J. H. (1998). Traumatic brain injury rehabilitation: Is it worth the cost? *Applied Neuropsychology, 5*, 184–193.
- Sunderland, M., Slade, T., & Andrews, G. (2012). Developing a short-form structured diagnostic interview for common mental disorders using signal detection theory. *International Journal of Methods in Psychiatric Research, 21*(4), 247–257.
- Wechsler, D. (1997). *WAIS-III administration and scoring manual*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1999). *WASI: Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.

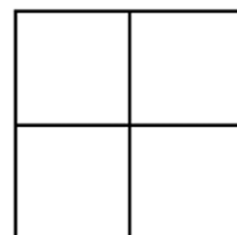
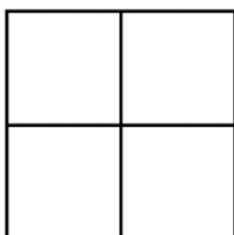
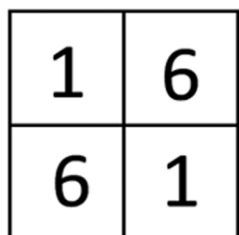
APPENDIX

DESIGN ORGANIZATION TEST

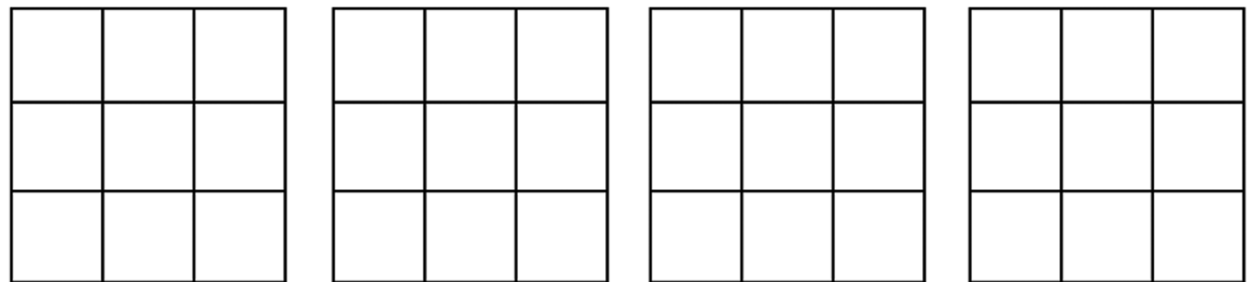
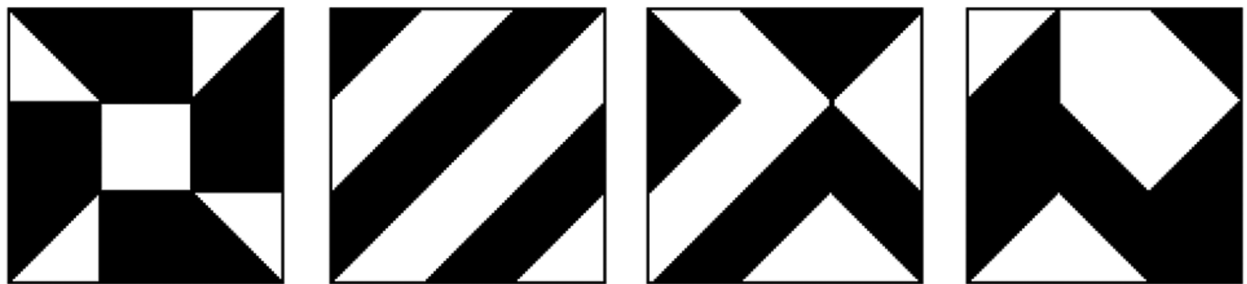
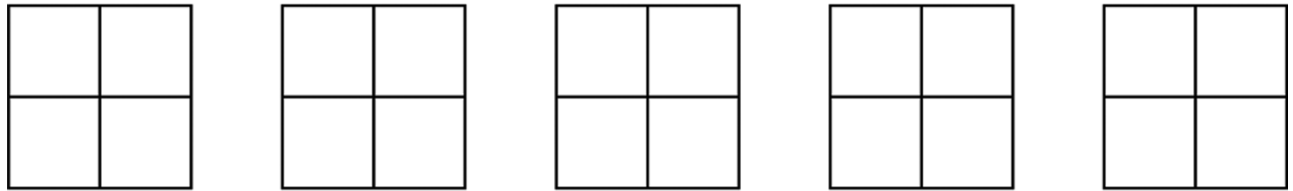
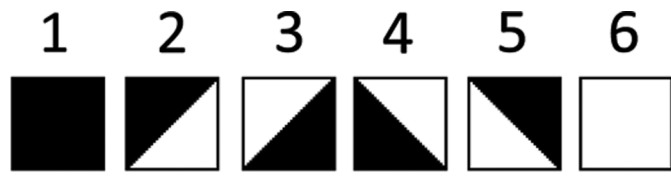
Practice



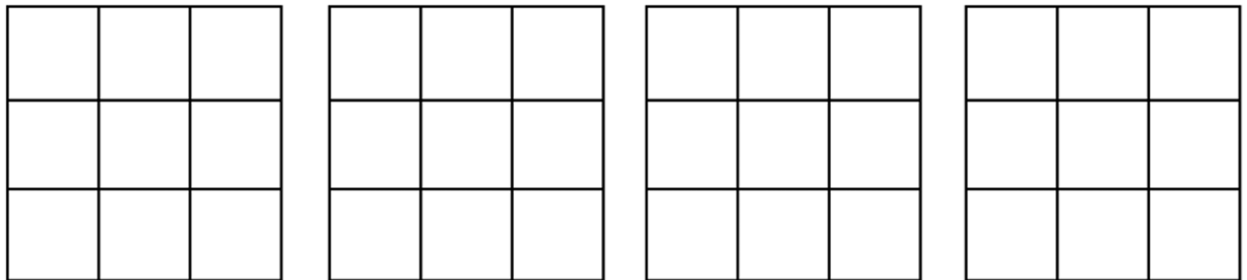
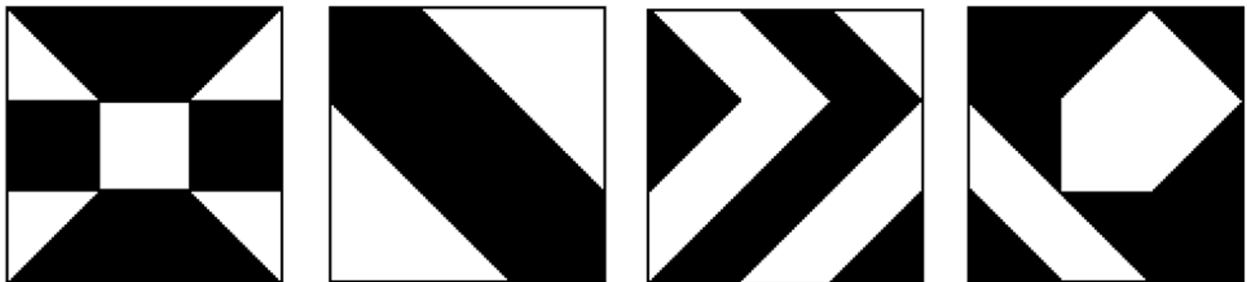
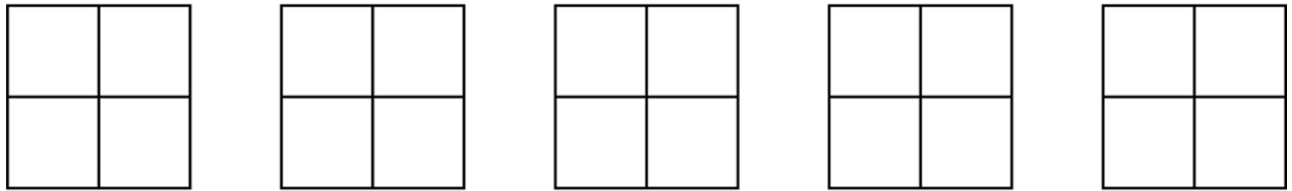
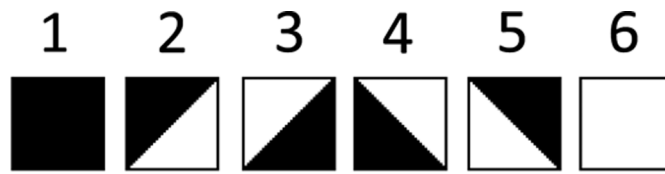
Example:



Form A



Form B



DOT Scoring Key - Form A

5	6	1	1	2	5	6	5	1	3
1	1	5	2	3	4	4	6	2	1

5	1	3	2	3	2	4	5	2	3	6	5
1	6	1	3	2	3	2	3	4	1	4	3
2	1	4	2	3	2	3	2	5	2	5	1

DOT Scoring Key - Form B

4	1	6	5	5	2	4	2	6	3
6	4	6	3	3	4	3	5	2	6

5	1	2	1	4	6	4	5	4	1	2	5
1	6	1	5	1	4	2	3	2	5	6	3
3	1	4	6	5	1	3	2	3	4	5	1