Personnel Assessment and Decisions

Volume 8 | Issue 1 Article 6

2022

The Attention to Detail Test: Measurement Precision and Validity Evidence for a Performance-Based Assessment of Attention to Detail

Brent A. Stevenor
Bowling Green State University

Michael John Zickar
Bowling Green State University - Main Campus

Fletcher Wimbush
The Hire Talent

Weston Beck

Recommended Citation

Stevenor, Brent A.; Zickar, Michael John; Wimbush, Fletcher; and Beck, Weston (2022) "The Attention to Detail Test: Measurement Precision and Validity Evidence for a Performance-Based Assessment of Attention to Detail," *Personnel Assessment and Decisions*: Number 8 : Iss. 1 , Article 6.
DOI: https://doi.org/10.25035/pad.2022.01.006
Available at: https://scholarworks.bgsu.edu/pad/vol8/iss1/6

The Attention to Detail Test: Measurement Precision and Validity Evidence for a Performance-Based Assessment of Attention to Detail

Brent A. Stevenor[1], Michael J. Zickar[1], Fletcher Wimbush[2], and Weston Beck[2]

1. Bowling Green State University
2. The Hire Talent

KEYWORDS

attention to detail, selection, item response theory

ABSTRACT

We report on the dimensionality, measurement precision, and validity of the Attention to Detail Test (ADT) designed to be a performance-based assessment of people's ability to pay attention to detail. Within the framework of item response theory, we found that a 3PL bifactor model produced the most accurate item parameter estimates. In a predictive validity study, we found that the ADT predicted supervisor ratings of subsequent overall job performance and performance on detail-oriented tasks. In a construct-related study, scores on the ADT correlated most strongly with the personality facet of perfectionism. The test also correlated with intelligence and self-reported ACT scores. The implications of modeling the ADT as unidimensional or multidimensional are discussed. Overall, our findings suggest that the ADT is a valid measure of attention to detail ability and a useful selection tool that organizations can use to select for detail-oriented jobs.

In many occupations it is important that employees pay close attention to detail to avoid making costly mistakes. For example, providing a patient with the wrong dosage of medication due to a misreading of the prescription or incorrectly recording the final digit of a high-profile client's phone number are mistakes due to a lack of attention to detail that have negative implications for organizations and their stakeholders. Organizations wishing to hire applicants who can pay attention to detail may require applicants to complete a self-report personality questionnaire, and applicants with high conscientiousness scores will be assumed to have high attention to detail ability. The problem is that assessing applicants' perceptions of their ability to pay attention to detail is not the same as assessing their actual ability. With self-report personality questionnaires, applicants can fake their scores to look more appealing to the hiring organization, which threatens the construct validity of the measure and thereby inhibits the inferences that can be made from the faking applicants' scores (Tett & Simonet, 2021). In fact, studies have shown that applicants may adjust their personality to appear as an ideal candidate whose personality closely aligns with the culture of the hiring organization (Canagasuriam & Roulin, 2021; Roulin & Krings, 2020). Therefore, applicants who apply for detail-oriented jobs may fake their responses to conscientiousness items to present themselves as high in attention to detail.

One solution to this problem is to assess applicants' *ability* to pay attention to detail. The purpose of this manuscript is to introduce the Attention to Detail Test (ADT), which is a performance-based assessment of attention to detail that can be used as a prehire assessment tool when making personnel selection decisions. Within the framework of item response theory (IRT), we provide evidence for the measurement precision and validity of the ADT. This assessment benefits research and practice such that it is a valid personnel selection tool that can be used to predict the future job performance of applicants applying to detail-oriented jobs.

Corresponding author:
Brent Stevenor
Author Email: basteve@bgsu.edu

Defining Attention to Detail

From a neuropsychological perspective, attention is defined as, “the regulating of various brain networks by attentional networks involved in maintaining the alert state, orienting, or regulation of conflict” (Posner & Rothbart, 2007, p. 2). There are three attention networks (i.e., alerting, orienting, and executive) that are located in various brain regions and are responsible for sensing/perceiving stimuli and resolving attention-oriented tasks. As Posner and Rothbart (2007) described in their review of research on attention networks, people differ in attentional ability, and these differences are attributed to biology as well as socialization and culture.

From an intelligence research perspective, the Cattell-Horn-Carroll (CHC) theory of cognitive abilities (McGrew, 2005) posits that there is a general factor of cognitive ability that reflects onto a variety of narrower abilities. Processing speed, or the ability to perform simple and repetitive cognitive tasks such as identifying if two pairs of information are the same or different, is a narrow component of the CHC model that falls under the broad ability of controlled attention (Schneider & Newman, 2015). As Schneider and Newman (2015) noted, *attentional fluency* is a more appropriate label for processing speed, as performance on assessments of this ability is determined by peoples’ ability to focus their attention on sequentially occurring stimuli. Typical assessments of attentional fluency are speeded, requiring test takers to quickly respond to a series of simple cognitive tasks, and the time it takes respondents to complete the test (i.e., mental speed) as well as the number of items correctly answered are recorded (e.g., Danthiir et al., 2005). Less common are tests that do not assess mental speed but do assess peoples’ ability to pay attention to detail. A benefit of nonspeeded attentional fluency tests is they allow for test takers to complete all test items to the best of their ability without making errors due to time pressure, whereas time pressure may cause undue errors on speeded tests or result in test takers not completing all the items. In this manuscript, we describe the ADT, which is a new, nonspeeded selection tool that assesses individual differences in attentional fluency.

Attention to detail has been conceptualized as an organizational cultural value (O’Reilly et al., 1991) as well as a narrow personality trait (Ashton & Lee, 2007; Hogan & Hogan, 2002). Organizations with attention to detail cultures are defined by high quality work, precision, compliance, and a low tolerance for mistakes (Miron-Spektor et al., 2007; O’Reilly et al., 1991). Studies have shown that positive outcomes are associated with having an attention to detail culture such that it can positively influence performance quality and productivity; however, overemphasizing attention to detail may inhibit innovation (Benner & Tushman, 2002; Naveh & Erez, 2004). During the selection process, organizations wishing to hire applicants that can pay attention to detail may administer a personality questionnaire that measures applicants’ perceptions of their conscientiousness and select applicants with the highest scores. As previously noted, the issue is that personality questionnaires do not assess actual ability to pay attention to detail, which warrants the need for a performance-based assessment of the construct.

From the personality perspective, attention to detail is considered a narrow personality trait that falls under the domain of conscientiousness. Conscientious people are described as thorough, organized, and precise in their work (John & Srivastrava, 1999). In their development of the Global Personality Inventory (GPI), Schmit et al. (2000) defined attention to detail as, “A desire for accuracy, neatness, thoroughness, and completeness; the ability to spot minor imperfections or errors; and a meticulous approach to performing tasks” (p. 185). In this manuscript, we introduce a performance-based assessment of attention to detail that organizations can use to differentiate between applicants in the selection process.

The Attention to Detail Test

Currently, organizations wishing to evaluate applicants’ attention to detail may use a self-report questionnaire that measures subjective evaluations of the construct. The issue with using personality questionnaires to measure this construct is they are prone to faking (Zickar & Drasgow, 1996), and they do not assess the ability to pay attention to detail; rather, they assess perceptions of ability to pay attention to detail. As a solution, we introduce the ADT, which is a 26-item, multiple-choice, performance-based assessment of attention to detail that organizations can use as a prehire assessment tool. The test contains three question types: (a) name and phone number, (b) email addresses, and (c) name and address. Each item has two columns of information, and participants are tasked with determining whether the information in the left column matches the information in the right column. Many of the items contain minor differences between the two columns (e.g., one letter or number is different), but some items contain identical information in both columns, and participants must determine whether the two columns differ. A sample item and instructions from the ADT are found in the appendix, and the full assessment is available at https://www.preemploymentassessments.com/short-detail-test/.

The ADT items are like those found in the Minnesota Clerical Test[1] (MCT; Andrew et al., 1979), which is a speeded, two-part test that contains 200 number and name checking items. The MCT was used as a selection tool for

1 The original title in 1933 was the *Minnesota Vocational Test for Clerical Workers*.

clerical positions and assessed applicants' verbal and numerical acuity. Like the ADT, applicants were required to indicate whether a pair of numbers or words was identical, and separate scores were given for numerical and verbal acuity. Clerical aptitude tests such as the MCT have been meta-analytically found to predict job proficiency, training success, and performance of clerical workers (Pearlman et al., 1980; Whetzel et al., 2011).

The ADT differs from the MCT in that the ADT is not speeded and contains many fewer items. This allows hiring personnel to compare applicants based on scores from the same number of items, and the ADT score is more indicative of attention to detail ability rather than test-taking speed. Also, some ADT items contain both numbers and words (e.g., the name and phone number items), and only a single score is given rather than separate numerical and verbal acuity scores. Additionally, whereas in the MCT applicants were tasked with determining if the numbers or words within a pair were identical, the ADT requires applicants to indicate if the two columns are identical and, if not, which rows the information differs on. We posit that requiring applicants to not only identify *if* a difference exists but also *where* is a more rigorous test of ability. Last, we reconceptualize the ADT to be a measure of general attention to detail ability, a broader construct than clerical skills ability. As described in the following studies, the ADT has been validated in samples that include clerical and nonclerical workers; therefore, the ADT can be considered for use as a selection tool for any job that consists of detail-oriented tasks (e.g., accountant). In the following sections, we present information on the dimensionality and measurement precision of the ADT and provide evidence for the convergent, discriminant, and predictive criterion-related validity of the test.

Item Analysis

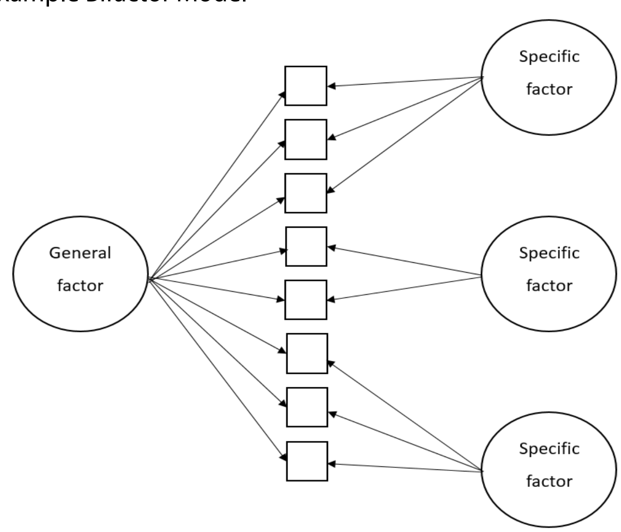
Participants, Procedure, and Analytical Strategy

The ADT was administered by The Hire Talent to 17,106 job applicants applying for positions such as accountants, receptionists, and bookkeepers. Of those who reported demographic information (n = 4,671), 66% were female, 12% African American, 1% American Indian or Alaska Native, 10% Asian, 14% Hispanic, 1% Native Hawaiian or Pacific Islander, and 55% White.

We used IRT to evaluate the dimensionality and measurement precision of the ADT. The benefits of using IRT for test development are that person and item parameters are simultaneously modeled (Embretson & Reise, 2013), and the measurement precision of each item along the ability continuum can be evaluated (Zickar, 1998). Using IRT, we were able to evaluate and determine whether each item precisely distinguished between people high and low in attention to detail ability. It is important to note that the ADT is not a speeded test, as IRT is not suitable for evaluating such tests.

The ADT is intended to assess a general ability to pay attention to detail, but being that it has three question types, we fit various unidimensional and multidimensional models to the data to examine the dimensionality of the test. Specifically, we compared the fit of six IRT models: unidimensional two-parameter logistic (2PL) and three-parameter logistic (3PL) models, three-factor 2PL and 3PL models, and bifactor 2PL and 3PL models. For the unidimensional models, all 26 items were fixed to load onto a single factor. For the three-factor models, the name and phone number items, email address items, and name and address items were fixed to load onto three separate factors. For the bifactor models, all 26 items were fixed to load onto a general ability factor and the three item types were also fixed to load onto three specific factors that were uncorrelated with the general factor (see Figure 1). Rather than only comparing whether the ADT data fit a one- or three-factor model, fitting a bifactor model allowed us to examine whether there is a general factor of attention to detail ability that is uncorrelated with the three specific factors that may appear due to item similarity rather than the existence of three specific abilities (Holzinger & Swineford, 1937). These specific factors are called *testlets*, or groups of items that have similar content (DeMars, 2012). With tests such as the ADT that have clusters of similar items, it is important to fit a bifactor model to the data to examine the influence of the testlets on responses to the items, as failing to do so may result in inaccurate item parameter estimates (DeMars, 2006).

FIGURE 1.
Example Bifactor Model



Unidimensional IRT models assume that the probability of answering an item correctly is dependent on a single ability factor, θ. Multidimensional IRT models assume that people use multiple abilities when responding to a single test item (Reckase, 2009), or as is the case with the ADT, these models allow researchers to partition out variance that is due to similarity in item content (rather than specific abilities) from the variance that is accounted for by the general ability factor that a test is measuring. Our reason for fitting various models to the ADT data is that it is important to specify the correct model so that parameter estimates can be accurate. In addition, testing different models can provide insight into the response process used by test takers.

The 2PL and 3PL are IRT models that can be fitted to unidimensional and multidimensional tests with dichotomously scored items. The unidimensional 2PL model has two parameters: location (*b*) and discrimination (*a*). The location (*b*) parameter indicates the point along the ability (θ) continuum that a person has a 50:50 chance of correctly endorsing an item. The discrimination (*a*) parameter indicates how well an item discriminates between people at *b*. The unidimensional 3PL model has three parameters: location (*b*), discrimination (*a*), and a pseudo-guessing parameter (*g*), which indicates the probability that a person with low ability will correctly endorse an item by guessing (Zickar, 1998). In multidimensional IRT, the location parameter is represented by *d*, and difficulty is represented by *D*, which is analogous to the *b* parameter from unidimensional models. A negative *D* value indicates an item is easy, whereas a positive *D* value indicates an item is difficult (Ackerman et al., 2003; Reckase, 2009). Additionally, the *a* parameter is analogous with a factor loading from a traditional exploratory factor analysis, and large values (i.e., $a \geq 1$) indicate the item effectively differentiates between people high and low in attention to detail ability (Zickar et al., 2002).

Results

For all IRT analyses presented in this manuscript, we used the "mirt" package (Chalmers, 2012) in R. The fit statistics for all six IRT models are found in Table 1. Relative to the other models, the 3PL bifactor model in which the general factor of attention to detail was not allowed to correlate with the three specific factors (i.e., testlets) fit the data the best[2] (-2LL = -151016.70, AIC = 302241.40, BIC = 303047.10, M2 = 1665.99, RMSEA = .02, CFI = .99). Given the similarity of majority of the model fit statistics between the 2PL and 3PL bifactor models, we also conducted a likelihood ratio test and found that the 3PL bifactor model was a significant improvement in model fit ($\chi 2(N = 17{,}106) = 134.82$, $p < .05$). The item fit statistics and parameter estimates for the 3PL bifactor model are found in Table 2. The results suggest that all 26 items fit the model well and load strongly onto the general factor being that all RMSEA values were less than .06 and all *a* values on the general factor were greater than 1. Additionally, the loadings on the specific factors were strong for majority of the items, suggesting that items of the same type clustered together due to similarity of content. Last, the results suggest that the items on the test are relatively easy, with *D* values ranging from -1.61 to -.48.

Because we modeled the ADT data with a bifactor structure, we used the "psych" package in R (Revelle, 2015) to calculate omega hierarchical (ωH) to estimate the proportion of variance in total ADT scores that is accounted for by the general factor (ωH = .72; McDonald, 1999). This estimate implies that 72% of the variance in total ADT scores is attributed to individual differences on the general factor. We also calculated omega hierarchical subscale (ωHS) for the "name and phone number" testlet (ωHS = .57), "email addresses" testlet (ωHS = .23), and "name and address" testlet (ωHS = .26), which is an estimate of the unique variance accounted for by each testlet once the general factor variance has been partitioned out (Reise et al., 2013). These estimates imply that the majority of the variance in total ADT scores is attributed to individual differences on the general factor; therefore, the ADT is essentially unidimensional (see Rodriguez et al., 2016 for a review on bifactor model statistical indices). The key takeaway from this item analysis is that the ADT measures a general factor of attention to detail ability, and the 26 items are not difficult but are effective at differentiating between people high and low in general ability to pay attention to detail. The utility of modeling the three testlets is explored further in the following studies.

Along with the item analysis, we conducted additional analyses to further evaluate the ADT. First, we calculated the proportion of test takers that achieved a perfect score. Out of 17,106 responses, only 2,410 (14.09%) applicants achieved a perfect score, providing further evidence that the ADT effectively differentiates between people of varying attention to detail ability despite its lack of a time limit. We then examined the relation between test duration and score to rule out the possibility that test score is a reflection of test-taking speed. We correlated ADT sum scores with duration and found that test score and duration of test time were not correlated ($r = .00$). We also examined the relation between item position and item difficulty to rule out the

2 When fitting bifactor models to data, it is important to demonstrate the invariance of the general factor across different sets of domains (see Eid et al., 2017 for a demonstration). In addition to the models presented in this manuscript, we fit three alternative bifactor-(S-1) models to the data and found strong evidence for the invariance of the ADT general factor. Please contact the first author for additional detail regarding these results.

3 The sample size for supervisor ratings of performance on detail-oriented tasks was $N = 320$, and $N = 177$ for ratings of overall job performance.

TABLE 1.
IRT Model Fit Statistics

Model	-2LL	AIC	BIC	M2	RMSEA	CFI
Unidimensional						
2PL	-163278.80	326661.90	327064.50	46769.58	.10	.94
3PL	-162005.80	324167.70	324772.00	22701.35	.07	.97
Multidimensional						
2PL three factor	-152487.10	305084.20	305510.30	7958.48	.04	.99
3PL three factor	-152439.00	305040.10	305667.60	4692.25	.03	.99
2PL bifactor	-151084.10	302324.20	302928.50	4462.27	.03	.99
3PL bifactor	-151016.70	302241.40	303047.10	1665.99	.02	.99

Note. N = 17,106. The M2 is a limited-information test statistic that is robust to Type I error. Smaller values indicate better model fit (Maydeu-Olivares & Joe, 2005). -2LL = -2 log likelihood, AIC = Akaike information criterion, BIC = Bayesian information criterion, RMSEA = root mean squared error of approximation, CFI = comparative fit index.

possibility of fatigue effects. Item position and IRT item difficulty were negatively correlated (r = -.47), suggesting that items at the end of the test were answered correctly more than items at the beginning (lower IRT difficulty parameters represent easier items). These results rule out fatigue effects as an alternative explanation. Last, we briefly examined whether there were race and gender differences in ADT scores. To examine race and gender differences, we created sum scores for each group and calculated Cohen's d. Due to unequal racial group sizes, we combined all non-White racial groups into a composite variable. As shown in Table 3, mean differences in ADT sum scores were negligible between White and non-White applicants (d = -.16), and between female and male applicants (d = .09).

Predictive Criterion-Related Validity

Participants, Procedure, and Analytical Strategy

To demonstrate the predictive criterion-related validity of the ADT, we correlated ADT scores of 320 job applicants who were hired after completing the ADT with supervisor performance ratings. The applicants included in this study were derived from the sample of N = 17,106 used for the item analysis. The average duration between ADT test administration and supervisor performance evaluations was 350.91 (SD = 288.18) days.

To score the ADT within the IRT framework, we fit the 3PL bifactor model to the data and calculated factor scores using the expected a posteriori method (Embretson & Reise, 2013). Factor scores are latent trait estimates that can fall above or below 0 and indicate whether a person is above or below average in ability. For example, a person with a factor score of 1 is estimated to be one standard deviation above the average in attention to detail ability. Being that we fit a 3PL bifactor model to the data, each respondent had four latent trait estimates: one for the general factor and one for each testlet. For comparison purposes, we also calculated sum scores on the ADT from the traditional CTT framework. We then correlated the ADT scores with supervisor ratings of overall job performance and performance on detail-oriented tasks, which ranged from 1 (low) to 10 (high).

Results

Descriptive statistics and intercorrelations for the ADT and supervisor performance ratings are found in Table 4. ADT general factor scores significantly predicted supervisor ratings of overall performance (r = .20, p < .05) and performance on detail-oriented tasks (r = .24, p < .05). ADT sum scores significantly predicted supervisor ratings of detail-oriented performance (r = .19, p < .05) but not overall performance (r = .12, *ns*). The three testlet factor scores did not significantly predict supervisor ratings of overall performance nor performance on detail-oriented tasks. The correlation between the ADT general factor scores and sum scores was .91, suggesting that scores produced by both methods are strongly related. It is important to note, however, that sum scoring the ADT rather than fitting a 3PL bifactor model resulted in weaker correlations between ADT scores and supervisor performance ratings. Although the testlets did not predict supervisor performance ratings, modeling the testlets resulted in stronger correlations between the ADT general factor scores and supervisor performance ratings.

Convergent and Discriminant Validity

Participants, Procedure, and Analytical Strategy

To demonstrate convergent and discriminant validity for the ADT, we examined correlations between ADT gen-

TABLE 2.
IRT Item Fit Statistics and Parameter Estimates for the ADT

	General	Specific 1	Specific 2	Specific 3				
Item	a_1	a_2	a_3	a_4	*d*	*g*	*D*	RMSEA
1	2.60	2.49	-	-	3.24	.00	-0.90	.03
2	2.65	2.50	-	-	1.74	.00	-0.48	.02
3	3.32	3.20	-	-	4.41	.00	-0.96	.03
4	12.18	12.84	-	-	11.64	.01	-0.66	.02
5	2.45	2.05	-	-	2.45	.00	-0.77	.02
6	3.87	3.63	-	-	3.01	.00	-0.57	.02
7	9.24	9.78	-	-	8.99	.01	-0.67	.02
8	3.10	2.87	-	-	2.08	.00	-0.49	.03
9	1.74	-	1.31	-	1.77	.00	-0.81	.01
10	4.66	-	0.17	-	6.49	.00	-1.39	.03
11	2.32	-	0.08	-	2.57	.00	-1.10	.02
12	2.90	-	0.06	-	4.05	.00	-1.40	.04
13	3.72	-	3.26	-	4.85	.00	-0.98	.02
14	3.78	-	3.03	-	4.79	.00	-0.99	.02
15	1.16	-	-0.04	-	0.65	.00	-0.56	.02
16	1.66	-	0.83	-	1.47	.00	-0.80	.02
17	1.79	-	0.38	-	0.88	.20	-0.48	.03
18	1.38	-	-	0.90	1.09	.02	-0.66	.02
19	1.73	-	-	0.81	2.58	.00	-1.35	.02
20	3.38	-	-	1.81	5.77	.00	-1.50	.04
21	1.42	-	-	1.11	2.15	.00	-1.19	.01
22	2.00	-	-	1.58	2.77	.00	-1.09	.01
23	2.40	-	-	1.31	4.41	.00	-1.61	.03
24	3.08	-	-	1.66	5.30	.00	-1.52	.03
25	1.30	-	-	0.97	1.25	.00	-0.77	.02
26	1.80	-	-	1.43	2.37	.00	-1.03	.01

Note. *N* = 17,106. General = general factor, Specific 1 = “Name and phone number” (testlet) factor, Specific 2 = “Email addresses” (testlet) factor, Specific 3 = “Name and address” (testlet) factor, *a* = discrimination parameter, *d* = location parameter, *g* = guessing parameter, *D* = difficulty parameter, RMSEA = root mean squared error of approximation.

eral factor scores and sum scores and a variety of external correlates. For convergent validity, we expected ADT scores to positively correlate most strongly with conscientiousness compared to the other Big 6 personality domains, and we also expected ADT scores to positively correlate most strongly with perfectionism, defined as a concern for detail, compared to the other facets of conscientiousness being that attention to detail is treated as a facet of conscientiousness in the personality literature (Ashton & Lee, 2007; Hogan & Hogan, 2002; Schmit et al., 2000). Attention to detail has also been shown to positively relate with performance (Muchinsky, 1993); therefore, we expected ADT scores to positively correlate with measures of academic performance (GPA and ACT scores) as well as with scores on the Sandia Matrices (Harris et al., 2020), which are a measure of intelligence that requires respondents to pay attention to

TABLE 3.
Number of Items Correct by Race and Gender

	M	*SD*
Race		
African American	19.22	6.90
American Indian or Alaska Native	20.93	6.06
Asian	20.70	6.69
Hispanic	20.04	6.22
Native Hawaiian or Pacific Islander	19.96	5.27
Non-White (composite)	20.02	6.52
White	21.03	5.93
Gender		
Female	20.89	5.96
Male	20.35	6.52

Note. Sum scores ranged from 0 to 26.

detail and select the option that best fits a pattern of shapes and colors. For discriminant validity, we expected ADT scores to have weak relations with the other five domains of the HEXACO. Given our assertion that the three testlets of the bifactor model exist due to item similarity rather than the existence of specific abilities, we did not expect scores on the specific factors to correlate with any of the external variables of interest; therefore, we only discuss correlations between the general factor and external correlates in the results.

A sample of 145 undergraduate psychology students from a medium-sized midwestern university was recruited to participate in this study; they received course extra credit for participating. Participants were removed from the data if they missed two out of two attention check items. All participants met our inclusion criteria, resulting in a final sample of $N = 145$ that was 19.55 ($SD = 1.45$) years old, 72% female, 77% White, and 17% African American.

Measures

Attention to Detail. The 26-item ADT was used to measure attention to detail.

Personality. The 100-item HEXACO-PI-R (Lee & Ashton, 2018) was used to measure personality. The HEXACO-PI-R is a measure of the Big 6 domains of personality as well as four facets per domain. Participants indicated their level of agreement with each item on a scale from 1 = *Strongly disagree* to 5 = *Strongly agree*.

Intelligence. Intelligence was measured using Harris et al.'s (2020) Sandia Matrices. The 10-item set of object relation and logic items was used in this study. For both item types, participants were tasked with selecting the option that best completed a pattern of images. The object relation items varied in shape, shading, or orientation, and the logic items varied in shape, size, and involved conjunction or disjunction (i.e., objects located on top of one another).

GPA. Participants self-reported their cumulative undergraduate GPA.

ACT. Participants self-reported their ACT score. If participants only completed the SAT, those scores were converted to ACT scores using the conversion calculator provided by Princeton Review (https://www.princetonreview.com/college-advice/act-to-sat-conversion).

Results

Descriptive statistics, reliability estimates, and intercorrelations for the variables included in this study are found in Table 5. Factor scores for the ADT were calculated by fitting a 3PL bifactor model and using the expected a posteriori method (Embretson & Reise, 2013), sum scores for the ADT were calculated by summing the number of items correctly answered, and average sum scores were calculated for the remainder of the scales used in this study. As predicted, ADT general factor ($r = .34$, $p < .05$) and sum scores ($r = .34$, $p < .05$) correlated most strongly with perfectionism compared to the other facets of conscientiousness. Additionally, ADT general factor and sum scores were significantly positively correlated with intelligence ($r = .32$,

TABLE 4.
Descriptive Statistics and Intercorrelations for ADT and Supervisor Performance Ratings

		M	*SD*	1	2	3
1.	ADT general factor	0.30	0.76	-		
2.	ADT sum score	22.39	4.90	.91	-	
3.	Overall performance	7.76	2.05	.20	.12	-
4.	Detail-oriented performance	7.88	2.13	.24	.19	.86

Note. Correlations with overall performance are based on a sample of N = 177, and correlations with detail-oriented performance are based on a sample of $N = 320$. All correlations are statistically significant ($p < .05$) except for the correlation between ADT sum score and overall performance.

TABLE 5.
Descriptive Statistics, Reliability, and Intercorrelations Between ADT and Correlates

	ADT (g)	ADT (sum)	*M*	*SD*	*α*
Honesty-Humility	.13	.16	3.39	0.56	.79
Sincerity	.03	.07	3.31	0.70	.52
Fairness	-.03	.02	3.48	0.87	.73
Greed avoidance	.10	.11	3.01	0.89	.74
Modesty	.29	.25	3.76	0.75	.66
Emotionality	.09	.11	3.54	0.60	.83
Fearfulness	.03	.04	3.31	0.75	.57
Anxiety	.15	.17	3.98	0.80	.69
Dependence	-.06	-.04	3.17	0.82	.65
Sentimentality	.16	.17	3.71	0.78	.69
Extraversion	-.11	-.09	3.19	0.62	.85
Social self-esteem	.01	.05	3.40	0.73	.56
Social boldness	-.16	-.16	2.82	0.84	.71
Sociability	-.11	-.09	3.34	0.83	.71
Liveliness	-.06	-.05	3.21	0.81	.73
Agreeableness	.00	-.03	3.00	0.50	.78
Forgiveness	-.26	-.30	2.52	0.76	.70
Gentleness	.13	.13	3.27	0.61	.42
Flexibility	-.01	-.03	3.02	0.70	.56
Patience	.16	.14	3.21	0.75	.67
Conscientiousness	.18	.21	3.45	0.55	.82
Organization	.06	.06	3.34	0.92	.76
Diligence	.09	.13	3.76	0.69	.68
Perfectionism	.34	.34	3.54	0.64	.52
Prudence	.07	.14	3.18	0.73	.65
Openness to Experience	.16	.17	3.23	0.54	.75
Aesthetic appreciation	.13	.15	3.31	0.78	.50
Inquisitiveness	-.04	-.04	2.77	0.85	.58
Creativity	.24	.24	3.44	0.82	.67
Unconventionality	.12	.14	3.39	0.60	.41
Intelligence	.32	.33	4.76	2.09	.61
GPA	.20	.18	3.34	0.58	
ACT	.30	.32	23.07	3.74	
Mean	0.01	21.72			
SD	0.66	3.81			

Note. N = 145. ADT (g) = General factor score from IRT 3PL bifactor model. ADT (sum) = Sum score calculated from traditional CTT perspective. Correlations greater than .22 are statistically significant ($p < .05$).

$p < .05$; $r = .33$, $p < .05$ respectively) and ACT scores ($r = .30$, $p < .05$; $r = .32$, $p < .05$ respectively). Contrary to our predictions, ADT general factor and sum scores did not significantly correlate with conscientiousness ($r = .18$; $r = .21$ respectively) nor GPA ($r = .20$; $r = .18$ respectively), although the pattern of the correlations was in the expected direction, the magnitudes were small to moderate, and the correlation between ADT and conscientiousness was the largest compared to the other five HEXACO domains. Aside from the expected relations, ADT general factor and sum scores correlated with the modesty facet of honesty-humility ($r = .29$, $p < .05$; $r = .25$, $p < .05$ respectively), the forgiveness facet of agreeableness ($r = -.26$, $p < .05$; $r = -.30$, $p < .05$ respectively), and the creativity facet of openness to experience ($r = .24$, $p < .05$; $r = .24$, $p < .05$ respectively).

DISCUSSION

The purpose of this manuscript is to evaluate the dimensionality of the ADT and present measurement precision and validity evidence. To provide accurate item parameter estimates, we first fit a series of IRT models to a large dataset of job applicant responses to the ADT and found that a 3PL bifactor model fit the data the best relative to other unidimensional and multidimensional models. The 3PL bifactor model that we fit to the data had one general ability factor and three specific factors, referred to as testlets, that we believe exist due to similarity in item content rather than the existence of three specific abilities that are unrelated to the general factor. The discrimination parameters (i.e., factor loadings) suggest that there is a dominant general factor of attention to detail ability and three testlets onto which the items load.

To determine whether modeling the testlets influenced the validity of the ADT, we scored the ADT from a multidimensional IRT approach, and for comparison purposes, we also scored the ADT from a unidimensional CTT approach. Specifically, we fit an IRT 3PL bifactor model to the data and calculated factor scores for the general factor and the three testlets. We also calculated a single sum score by adding up the number of items correctly answered for each job applicant.

In the predictive criterion-related validity study, we found that the ADT general factor score correlated more strongly with supervisor performance ratings compared to the ADT sum score. Although the testlets did not predict supervisor performance ratings, modeling the testlets resulted in larger correlations between the ADT general factor and performance, likely due to controlling for irrelevant method variance. We did find, however, that the ADT general factor and sum score were strongly correlated. In the convergent and discriminant validity study, correlations between the ADT general factor and sum score with the external constructs were consistently similar, suggesting that sum scoring the ADT is sufficient. Overall, we recommend fitting a 3PL bifactor model to ADT data to achieve accurate item parameter estimates and stronger criterion-related validity coefficients. We do recognize, however, the convenience of treating the ADT as unidimensional and calculating a single sum score. In addition, fitting overly complex models to data collected from smaller samples may result in increased error by capitalizing on chance (DeMars, 2012). Therefore, practitioners wishing to select applicants based on their ADT sum scores should feel comfortable doing so, as we demonstrated that the ADT is essentially unidimensional and a valid predictor of supervisor performance ratings. We do recommend, however, that in order to achieve the most accurate parameter estimates and validity coefficients that a bifactor model be fit to the data. If a 3PL bifactor model is fit to the ADT data, testlet scores should not be considered when evaluating applicants' attention to detail ability, as our results suggest that the testlets are necessary for achieving accurate parameter estimates but are not valid predictors of supervisor ratings of job performance.

After examining the dimensionality of the ADT, we conducted an item analysis and found that all 26 items are relatively easy yet effectively distinguish between people high and low in attention to detail ability. Additionally, we examined differences in sum scores across race and gender, and found negligible effect size differences between White and non-White applicants as well as between women and men, suggesting the ADT is not biased toward a specific race or gender. Then, we examined the test's criterion-related validity and found that scores on the ADT significantly predicted supervisor ratings of overall job performance and performance on detail-oriented tasks. Last, we examined the test's convergent and discriminant validity and found that ADT scores significantly correlated with perfectionism, intelligence, and self-reported ACT scores. The ADT also correlated with modesty and forgiveness; therefore, future research should continue to examine the discriminant validity of the test. Overall, the results suggest that the ADT is a valid performance-based assessment of attention to detail that researchers and practitioners could use to predict the future job performance of applicants.

Limitations and Future Directions

There are certain limitations to our examination of the ADT that should be considered and addressed in future studies. First, our sample for the item analysis was large and consisted of actual job applicants rather than participants recruited from an online crowdsourcing platform, but we were able to attain demographic information for only a portion of the applicants. Due to this limitation, we focused our assessment of adverse impact on mean differences at the scale level. Future research should examine measurement equivalence at the item-level across gender and race using CFA and IRT. We encourage future research to test if any ADT items demonstrate differential item functioning or if people from all genders and races have equal probabilities of answering the items correctly.

A second limitation is that we did not examine incremental validity for the ADT. Although we demonstrated evidence of criterion-related validity, future studies should examine whether the ADT predicts job performance over and above other common predictors (i.e., cognitive ability, conscientiousness, etc.). Previous meta-analyses by Barrick and Mount (1991) and Zettler et al. (2020) reported meta-analytic correlations of .22 and .28 between conscientiousness and job performance, respectively. Additionally, Harari et al. (2018) reported a meta-analytic correlation of .02 between perfectionism and job performance. Based on our finding in this manuscript that ADT scores had a small-to-moderate correlation with conscientiousness, it is plausible that the ADT predicts unique variance in job performance for which traditional personality measures do not account. The small-to-moderate correlation between conscientiousness and ADT scores also suggests that high self-report conscientiousness scores do not necessarily imply that a candidate has high attention to detail ability; therefore, organizations wishing to select candidates with high attention to detail ability will benefit from using the ADT in addition to a traditional self-report conscientiousness scale. Our goal in this manuscript was to provide initial validity evidence for the ADT, but future research should continue to examine the test's predictive ability.

Conclusion

Predicting the performance of job applicants who are applying to detail-oriented jobs had previously relied on self-report personality questionnaires that assess conscientiousness and its relevant facets. In this manuscript, we introduced the ADT as a performance-based alternative to assessing attention to detail ability. Our results suggest that the ADT is a valid predictor of job performance that precisely distinguishes between people high and low in attention to detail ability. The ADT can serve as a useful selection tool for organizations wishing to hire detail-oriented applicants.

REFERENCES

Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. Educational Measurement: Issues and Practice, 22, 37–51. https://doi.org/10.1111/j.1745-3992.2003.tb00136.x

Andrew, D. M., Paterson, D. G., & Longstaff, H. P. (1979). Minnesota clerical test. Psychological Corporation, Harcourt Brace & Company.

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. Personality and Social Psychology Review, 11, 150–166. https://doi.org/10.1177/1088868306294907

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. Personnel Psychology, 44, 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Benner, M. J., & Tushman, M. (2002). Process management and technological innovation: A longitudinal study of the photography and paint industries. Administrative Science Quarterly, 47, 676–707. https://doi.org/10.2307/3094913

Canagasuriam, D., & Roulin, N. (2021). The effect of organizational culture on faking in the job interview. Personnel Assessment and Decisions, 7. https://scholarworks.bgsu.edu/pad/vol7/iss1/8

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. Journal of Statistical Software, 48, 1–29. https://doi.org/10.18637/jss.v048.i06

Danthiir, V., Wilhelm, O., Schulze, R., & Roberts, R. D. (2005). Factor structure and validity of paper-and-pencil measures of mental speed: Evidence for a higher-order model? Intelligence, 33, 491–514. https://doi.org/10.1016/j.intell.2005.03.003

DeMars, C. E. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. Journal of Educational Measurement, 43, 145–168. https://doi.org/10.1111/j.1745-3984.2006.00010.x

DeMars, C. E. (2012). Confirming testlet effects. Applied Psychological Measurement, 36, 104–121. https://doi.org/10.1177/0146621612437403

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. Psychological Methods, 22, 541–562. https://doi.org/10.1037/met0000083

Embretson, S. E., & Reise, S. P. (2013). Item response theory. Psychology Press.

Harari, D., Swider, B. W., Steed, L. B., & Breidenthal, A. P. (2018). Is perfect good? A meta-analysis of perfectionism in the workplace. Journal of Applied Psychology, 103, 1121-1144. https://doi.org/10.1037/apl0000324

Harris, A., McMillan, J., Listyg, B., Matzen, L., & Carter, N. (2020). Measuring intelligence with the sandia matrices: Psychometric review and recommendations for free raven-like item sets. Personnel Assessment and Decisions, 6. https://scholarworks.bgsu.edu/pad/vol6/iss3/6

Hogan, R., & Hogan, J. (2002). Hogan Personality Inventory manual. Hogan Assessment Systems.

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. Psychometrika, 2, 41–54. https://doi.org/10.1007/BF02287965

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. Handbook of personality: Theory and research, 2, 102-138.

Lee, K., & Ashton, M. C. (2018). Psychometric properties of the HEXACO-100. Assessment, 25, 543–556. https://doi.org/10.1177/1073191116659134

Maydeu-Olivares, A., & Joe, H. (2005). Limited- and full-information estimation and goodness-of-fit testing in 2n contingency tables. Journal of the American Statistical Association, 100, 1009–1020. https://doi.

org/10.1198/016214504000002069
McDonald, R. P. (1999). Test theory: A unified approach. Lawrence Erlbaum Associates Publishers.
McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities: Past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), Contemporary intellectual assessment: Theories, tests, and issues (pp. 136–181). Guilford Press.
Miron-Spektor, E., Erez, M., & Naveh, E. (2007). Balancing innovation attention-to-detail and outcome-orientation to enhance innovative performance. Academy of Management Proceedings, 2007, 1–7. https://doi.org/10.5465/ambpp.2007.26508076
Muchinsky, P. M. (1993). Validation of personality constructs for the selection of insurance industry employees. Journal of Business and Psychology, 7, 475–482. https://doi.org/10.1007/BF01013760
Naveh, E., & Erez, M. (2004). Innovation and attention to detail in the quality improvement paradigm. Management Science, 50, 1576–1586. https://doi.org/10.1287/mnsc.1040.0272
O'Reilly, C. A., Chatman, J., & Caldwell, D. F. (1991). People and organizational culture: A profile comparison approach to assessing person-organization fit. Academy of Management Journal, 34, 487–516. https://doi.org/10.5465/256404
Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training success in clerical occupations. Journal of Applied Psychology, 65, 373-406. https://doi.org/10.1037/0021-9010.65.4.373
Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. Annual Review of Psychology, 58, 1-23. https://doi.org/10.1146/annurev.psych.58.110405.085516
Reckase, M. D. (2009). Multidimensional item response theory. Springer.
Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. Journal of personality assessment, 95, 129-140. https://doi.org/10.1080/00223891.2012.725437
Revelle, W. (2015). Package "psych." Retrieved from http://cran.r-project.org/web/packages/psych/psych.pdf.
Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. Psychological Methods, 21, 137-150. https://doi.org/10.1037/met0000045
Roulin, N., & Krings, F. (2020). Faking to fit in: Applicants' response strategies to match organizational culture. Journal of Applied Psychology, 105, 130–145. https://doi.org/10.1037/apl0000431
Schmit, M. J., Kihm, J. A., & Robie, C. (2000). Development of a global measure of personality. Personnel Psychology, 53, 153–193. https://doi.org/10.1111/j.1744-6570.2000.tb00198.x
Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. Human Resource Management Review, 25, 12–27. https://doi.org/10.1016/j.hrmr.2014.09.004
Tett, R., & Simonet, D. (2021). Applicant faking on personality tests: Good or bad and why should we care? Personnel Assessment and Decisions, 7. https://scholarworks.bgsu.edu/pad/vol7/iss1/2
Whetzel, D. L., McCloy, R. A., Hooper, A., Russell, T. L., Waters, S. D., Campbell, W. J., & Ramos, R. A. (2011). Meta-analysis of clerical performance predictors: Still stable after all these years. International Journal of Selection and Assessment, 19, 41–50. https://doi.org/10.1111/j.1468-2389.2010.00533.x
Zettler, I., Thielmann, I., Hilbig, B. E., & Moshagen, M. (2020). The nomological net of the HEXACO model of personality: A large-scale meta-analytic investigation. Perspectives on Psychological Science, 15, 723–760. https://doi.org/10.1177/1745691619895036
Zickar, M. J. (1998). Modeling item-level data with item response theory. Current Directions in Psychological Science, 7, 104–109. https://doi.org/10.1111/1467-8721.ep10774739
Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. Applied Psychological Measurement, 20, 71–87. https://doi.org/10.1177/014662169602000107
Zickar, M. J., Russell, S. S., Smith, C. S., Bohle, P., & Tilley, A. J. (2002). Evaluating two morningness scales with item response theory. Personality and Individual Differences, 33, 11–24. https://doi.org/10.1016/S0191-8869(01)00131-3

RECEIVED 08/03/2021 ACCEPTED 12/07/2021

Appendix

Attention to Detail Test Example Item

DIRECTIONS: These comparisons consist of names and phone numbers. Compare the left sample to the one on the right. Both sides should match exactly. If they don't match:

just in the name, select A.
just in the phone number, select B.
in both the name and the phone number, select C.
in neither the name nor the phone number, select D.

Left	Right	Select the correct answer.
Martin Cannon	Martin Cannan	
677-4413	677-4413	A , B , C , D