

University of New Hampshire

## University of New Hampshire Scholars' Repository

---

Psychology Scholarship

College of Liberal Arts (COLA)

---

6-1-2017

### A closer look at the Test of Personal Intelligence (TOPI)

John D. Mayer

*University of New Hampshire, Durham, jack.mayer@unh.edu*

A. T. Panter

*University of North Carolina*

David R. Caruso

*Yale University*

Follow this and additional works at: [https://scholars.unh.edu/psych\\_facpub](https://scholars.unh.edu/psych_facpub)

---

#### Recommended Citation

Mayer, J.D., Panter, A. T., & Caruso (2017). A closer look at the Test of Personal Intelligence (TOPI). *Personality and Individual Differences*, 111, 301-311. DOI 10.1016/j.paid.2017.02.008

This Article is brought to you for free and open access by the College of Liberal Arts (COLA) at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Psychology Scholarship by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact [Scholarly.Communication@unh.edu](mailto:Scholarly.Communication@unh.edu).

Running Head: PERSONAL INTELLIGENCE

## A Closer Look at the Test of Personal Intelligence, Version 1.4

John D. Mayer

Department of Psychology, University of New Hampshire

A. T. Panter

L.L. Thurstone Psychometric Laboratory, Department of Psychology and Neuroscience,  
The University of North Carolina at Chapel Hill

David R. Caruso

Yale College Dean's Office, Yale University

### Author Notes

The authors discussed statistical issues frequently during the paper's preparation; we were grateful for additional perspectives on the analyses provided by Michelle Langer, Senior Statistician of the *American Institutes for Research*. The analyses were conducted both by the authors and by Dr. Langer, who served as an independent statistical consultant. We also are thankful for the comments of Dr. Langer and Dr. Kateryna Sylaska on earlier versions of this work.

Correspondence regarding this manuscript may be directed to any of the authors. John D. Mayer, Department of Psychology, University of New Hampshire, Durham, NH 03824. A. T. Panter, L. L. Thurstone Psychometric Laboratory and the Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, NC, 27599, and David R. Caruso, Yale College Dean's Office, Yale University, New Haven, CT, 06520.

### Abstract

Personal intelligence involves the capacity to reason about personality and personality-related information. In Study 1 ( $N = 10,318$ ), we conduct an item-level analysis of the *Test of Personal Intelligence* (TOPI) to explore people's problem-solving abilities in the area. Personal intelligence divided into Consistency and Dynamic Reasoning factors, a finding we cross-validated in Study 2 ( $N = 8,459$ ). In Study 3 ( $N = 384$ ), we reanalyzed previously-collected data to examine correlates of the two factors. Studying personal intelligence with ability-based measures of the concept creates a virtuous cycle of better measurement and better understanding of the ability.

KEY WORDS: Personal intelligence, intelligence, personality, Test of Personal Intelligence

*Prepublication version of:*

Mayer, J.D., Panter, A. T., & Caruso (2017). A closer look at the *Test of Personal Intelligence* (TOPI). *Personality and Individual Differences*, *111*, 301-311. DOI 10.1016/j.paid.2017.02.008

## A Closer Look at the Test of Personal Intelligence, Version 1.4

Personal intelligence is a recently-described mental ability that involves the capacity to reason about personality and personality-related information. According to the theory, people employ their knowledge of personality to understand both themselves and the personalities of others (Mayer, 2008). More specifically, people use their personal intelligence to solve problems in four areas: They (a) recognize personality-relevant information in themselves and others, “reading” people’s traits. They (b) use such information to form models of people’s personalities so as to understand themselves and others; they (c) guide their own and others’ choices making use of personality-relevant information to choose goals consistent with their interests and values. And lastly, they (d) systematize their goals and plans to to enhance the likelihood of achieving their aims (Mayer, 2008). Personal intelligence also has a foundation in our human heritage: Among our evolutionary ancestors, those who could better understand themselves and the people around them likely experienced adaptive advantages relative to others (cf., Buss, 2008; Dunbar, 2009).

Personality can be characterized as “the specific mental organization and processes that produce an individual’s characteristic patterns of behavior and experience” (DeYoung, 2015, p. 33). The personality system manages mental processes such as an individual’s motives and emotions, knowledge and intelligences, and awareness and self-control (DeYoung, 2015; Larsen & Buss, 2014; Von Bertalanffy, 1951). Personal intelligence enhances the self-guidance of personality by allowing for reasoning both about one’s own personality and the personalities of others (Mayer, 2014).

Certain often-studied intelligences of today, including quantitative and spatial intelligences, emphasize reasoning about aspects of things: the manipulation of numbers, or the rotation of objects in space; these intelligences are sometimes referred to as “cool” because they are relatively impersonal in the topics they concern (Mayer, Caruso, & Salovey, 2016). Other intelligences emphasize reasoning about aspects of people, such as—in the case of emotional intelligence measured as an ability—reasoning specifically about their emotions; these are sometimes referred to as “hot” because they often concern matters of personal concern. Personal intelligence is a further example of a person-focused intelligence, differing from emotional intelligence in that it concerns a far wider range of reasoning than specifically emotion-centered problem-solving (Mayer et al., 2016).

To explore whether personal intelligence existed and could be objectively measured, the present authors developed a *Test of Personal Intelligence (TOPI)* that consisted of four broad areas of questions that corresponded to the four areas of problem solving described by the theory, and that yielded scores keyed to each of those areas. The initial studies of the TOPI indicated that ability-based items about personality could be written, correct answers identified, and that people exhibited reliable individual differences in their reasoning capacities in the area (Mayer, Panter, & Caruso, 2012). The TOPI exhibited promising relationships with other variables as well: Personal intelligence exhibits moderate relations with measures of verbal, quantitative and spatial intelligences in the  $r = .17$  to  $.30$  range (Mayer & Skimmyhorn, in press), and correlates about  $r = .65$  with strategic emotional intelligence (i.e., understanding and managing emotions). Among the Big Five personality traits, it correlates with self-reported openness-closedness at approximately  $r = .15$  (Mayer & Skimmyhorn, in press; Mayer et al., 2012). Finally, the intelligence predicts consequential academic and performance outcomes incrementally above

measures of general intelligence (*g*) alone, suggesting its practical usefulness in high-stakes testing (Mayer & Skimmyhorn, in press).

The first confirmatory factor model of the TOPI employed scores keyed to the four problem-solving areas of the TOPI, and indicated that personal intelligence could be regarded as a single, unitary ability. At the same time, the four test sections overlapped more in content than was optimal for indicators of the construct. That initial confirmatory factor model “represented an...imperfect, first representation of the results,” that would require follow-up (Mayer et al., 2012, p. 136). The first look at the TOPI was, in a sense, a promissory note: Surely more could be learned about the structure of abilities in the area.

In this article, we provide an *itemized* payment on that promissory note in the form of an item-level reanalysis of the TOPI to better answer, among other questions, what abilities people bring to bear when problem-solving in the area. We regard any such abilities as likely to be distinct from the four *content* areas of problem solving questions (Mayer et al., 2016). As an analogy, people who study literature learn to understand both fiction and nonfiction. However, their ability to understand those texts draws on mental abilities that divide differently from the subject matter: Both kinds of texts require abilities in vocabulary skill and sentence comprehension. Identifying the specific abilities of personal intelligence is important to describing what the reasoning involves and providing education in the area.

A second purpose of the study is to learn more about how people’s abilities in personal intelligence are distributed across samples and the population. General intelligence is often considered to be normally distributed. As nutrition improves worldwide and positive cognitive stimulation increases, there is some evidence that general intelligence is becoming more negatively skewed (i.e., the majority of people score better), but the shape still is predominantly normal (e.g., Colom, Lluís-Font, & Andrés-Pueyo, 2005).

There is, however, a suggestion that the distribution for intelligences focused on people is considerably different (cf., Maul, 2012). This is theoretically important because it may reflect that at the high end of reasoning about people there are, in fact, few reliable rules by which to judge people and to predict their behaviors past a certain level of complexity. At the lower end of the distribution, a negatively-skewed tail of the intelligences may also reflect the existence of biopsychological conditions such as Autism-spectrum disorders (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001).

Finally, we examine the TOPI in relation to self-estimated personal intelligence, showing that ability in personal intelligence and self-judgment are distinct constructs, and we will examine a newly revised test in relation to its earlier versions—re-computing relationships between new test scales and earlier-studied criteria.

Studying a concept together with the way it is measured creates a virtuous cycle of better measurement and better understanding of the construct (Borsboom, Mellenbergh, & van Heerden, 2004; Hood, 2009; O’Sullivan, 2007). Test subscales are better justified when they are based on the actual structure of abilities measured by the test (Reise, Moore, & Haviland, 2010; Sinharay, Puhon, & Haberman, 2011). Moreover, understanding the abilities people use to solve problems in the domain can be helpful to providing education in an area.

## **The TOPI Series of Tests**

### **General Overview of the TOPI Tests**

The Test of Personal Intelligence (TOPI) was introduced in 2012 to help determine the existence of a personal intelligence; evidence accrued for the construct as the scale underwent several revisions from versions from 1.0 to 1.2 (Mayer et al., 2012). The present version, 1.4, is a 93-item subset of the TOPI 1.2 (Mayer, Panter, & Caruso, 2014). Most key characteristics of the TOPI have remained the same over forms: The test is divided into four areas of problem solving, reflecting the four problem-solving areas of the theory, from “Recognizing Information,” to “Systematizing Plans.”

The four problem solving areas of the TOPI 1.4 are overall divided into 13 clusters of more specific test items. Items in each cluster have a similar format to one another. Across the test, all items are multiple choice and each one has four alternatives. For example, the Recognizing Information area includes a “Recognizing inner motives” cluster, in which test-takers are asked to assess a person’s wants and needs and, from those, forecast a likely behavioral pattern. The first item of that group asks:

1. If a person wants to be with one or more people, talk to them, go out with them, and have a good time, the person is likely going to:
  - a. be in love
  - b. express warmth toward someone
  - c. meet a goal of excellence
  - d. socialize

The test-taker who answers this item correctly must assess the given behaviors of being with people, going out with them, and consequent enjoyment, and extract from them the motive most likely to have directed a person: in this case the need to socialize (alternative “d”). The logic required for each task is different. For example, a cluster in the Systematizing Plans area includes items that ask whether test-takers can recognize goals that conflict (“Problematic goals”); see Mayer et al. (2012) for examples of items from each section of the test.

The TOPI’s coverage of the areas helps ensure both that all included test questions have to do with key aspects of personal intelligence and that a broad range of relevant content areas are sampled, providing evidence for its appropriate coverage of content as recommended in the *Standards for Educational and Psychological Testing* (Joint\_Committee, 2014).

### **Veridical Scoring and Item Difficulty**

Correct answers to the TOPI are scored correct (1 point) or incorrect (0 points). Correct answers were identified with reference to relevant published research in the field of personality psychology based on generally-agreed upon phenomena in personality psychology. Personality research is complex and does not always lead to a consensus; the most agreed-upon findings are those that often are widely known, making for relatively easy test items. The trade-off, however, was the relative confidence that we were measuring valid knowledge of personality as opposed to, say, agreement with one theorist or theoretical school at the expense of another.

### **Self-Judged Personal Intelligence**

People’s personal intelligence also can be assessed with self-judgment items keyed to the theory, and we have created such a scale with items of the form “I don’t know who I am” (reverse scored), and “I read people’s intentions well” (answered from “strongly disagree” to “strongly agree” on a 5-point scale), called the Self-Estimated Personal Intelligence inventory (SEPI). Like most instruments measuring self-reported intelligence, the SEPI is likely saturated with construct-irrelevant variance including endorsements based on high (or low) self-esteem and reflecting respondents’ imperfect understanding of what good reasoning consists of. Self-reports of intelligence generally correlate  $r = .20$  or lower with actual assessments of mental

ability (Brackett & Mayer, 2003; Paulhus, Lysy, & Yik, 1998) and we expect similar findings here.

### Study 1. Large Sample Data Analysis of the TOPI 1.4R

#### Purposes and Hypotheses of the Study

The key purpose of the study was to identify any distinguishable abilities that make up personal intelligence—and should they exist—to describe such abilities, and to develop scales that could adequately measure those distinct areas of reasoning. We were guided in those aims by several hypotheses.

**Hypothesis 1.** The first hypothesis provided a key foundation for the item analyses: to ensure that participants were paying adequate attention to the items in the second half of the test (the TOPI items are administered in a fixed order) given that TOPI items may place a substantial cognitive load on respondents. To test Hypothesis 1, we examined participants' drop-offs in attention by examining their rates of endorsement of unlikely responses, as well as their overuse of a single letter response (long-string responding) over the length of the test. If there were a drop-off from the first to second halves, we further planned to screen out any participants who exhibited precipitous in that regard before conducting further analyses.

**Hypothesis 2: *Personal intelligence can be divided into two or more correlated factors.*** Our second hypothesis was that personal intelligence could be divided into two or more (probably) highly correlated factors in a simple structure factor model. Given that “broad” intelligences such as spatial, verbal and quantitative intelligences often correlate in the range of  $r = .70$  to  $.95$ , we expected that any subsidiary factors of personal intelligence would correlate between  $r = .75$  to  $.95$  as well. To estimate the practical significance of any factors beyond the first, we planned to estimate each factors' explained common variance (ECV) in the context of an alternative bifactor model, which uniquely allows for such estimates (Reise, 2012). We planned to retain any factor with less than about 85% of its variance explained by the overall construct, consistent with contemporary practice (e.g., O'Connor Quinn, 2014).

**Hypothesis 3. *Similar scale performance for men and women.*** We hypothesized that any factors obtained would be consistent across groups of both women and men (i.e., configural invariance). We also expected that women would score somewhat higher than men on the test as they do for people-centered intelligences more generally (Mayer, Salovey, & Caruso, 2002) and that the item discrimination and difficulty levels of individual items (controlling for the overall difference) would be similar for women and men.

**Hypothesis 4: *Better discrimination at low levels of ability.*** Previous research suggests that most people exhibit good skills at understanding one another in the realm of people-oriented, *hot* intelligences such as personal and emotional intelligences (see, for example, Fiori et al., 2014; Mayer et al., 2016). Although most people perform well, we expect a substantial number of people will nonetheless exhibit poorer, low performance as indicated both by a negative skew among TOPI scores and the test's more accurate discrimination of test-takers at low levels of ability.

**Hypothesis 5: *Reliable measurement.*** We expected that TOPI 1.4 scales would exhibit reliable measurement as indicated by coefficient alpha reliabilities, reliabilities in IRT theory, and reasonable standard errors of measurement.

**Hypothesis 6. *Uniqueness of the measure given the cognitive response process it elicits.*** We expected that scales of the TOPI 1.4 would correlate  $r = .10$  to  $.20$  range with a measure of Self-Estimated Personal Intelligence (SEPI).

## Method

### Participants

Participants were 10,618 test-takers drawn from seven samples, mostly from the United States military. The sample included 8,049 men, 2,261 women, and 6 other; the mean age of the sample was 21.09. Further details of the samples and procedures used to collect the data can be found in the footnotes to Table 1.

A master data file was assembled from these seven ethnically diverse samples, tested by the United States *Office of Economic and Manpower Analysis* (OEMA) between the fall 2012 and June 2015. The data file began with a sample of 1,114 individuals tested at the U.S. Military Academy at West Point and concluded with a composite sample of individuals tested while in in Officer Candidate School in June 2015. We chose June 2015 as our stopping point for inclusion because our samples contained over 10,000 participants by then and data delivered by the OEMA often slowed in the summer. All participants received the TOPI 1.4 with the exception of Group 1 who received a version of the TOPI 1.2 that contained all the items on the TOPI 1.4.

### Measures

Our measures included the 93-item Test of Personal Intelligence, Version 1.4 and a 16-item scale of Self-Estimated Personal Intelligence (SEPI-16); both scales were described earlier in the introduction.

## Results

### Overview

We begin by describing our process for screening participants, move to our factor analysis of the TOPI, an examination of two newly-formed factor-based scales using the techniques of item response theory, and then present statistics for the new scales we developed.

### Fall-Off in Attention and Screening Participants Based on Their Attention Levels (Hypothesis 1)

All TOPI data provided by the OEMA and in the Study 1 Archive had been prescreened to ensure all participants' responses were complete. Sample 6, from our laboratory, was also screened for complete responding, and for signs of haphazard responding (Lortie, 2015). To determine whether the remaining test-takers exhibited a drop-off of attention, we developed two scales of inattentive responding (Huang, Curran, Keeney, Poposki, & DeShon, 2012).

The first, Infrequency, counted the number of times a participant selected the *least endorsed*, most obviously incorrect, answer for each of the 93 TOPI items among the 10,618 respondents. Infrequency scales for the first and second halves of the test correlated  $r = .63$ . The second scale, Letter-Repetition, was set equal to the maximum number of times a respondent chose any of the four letter alternatives A, B, C, or D across items. Letter-repetition in the first and second halves of the test correlated  $r = .36$ . The Infrequency and Letter Repetition Scales for the whole test correlated  $r = .34$ .

**Attention fall-off.** Both the infrequency and letter-frequency scales exhibited substantial positive skew, indicating that most participants were responding meaningfully throughout. When we employed a nonparametric one-sample Kolmogorov-Smirnov analysis of the mean difference in scales across the first- and second test halves, however, we found evidence for more infrequent responding and more letter repetitions during the second half of the test ( $ps < .0001$ ).

Table 1

Overview of the Original Samples of Study 1 after Data Exclusions  $N = 10,318$ 

Sample Number and Description <sup>a</sup>	<i>N</i> Total	Sample size <i>N</i> s after Screening		Age		
		Total	Men	Women	Mean	Range
United States Military Academy Cadets, West Point, Class of 2014, testing by the Office of Economic and Manpower Analysis <sup>b</sup>	1114	1106	905	201	20.98	20-30
As in sample 1, Class of 2015 <sup>b</sup>	1048	1048	867	174	20	20-22
As in sample 1, Class of 2016 <sup>b</sup>	972	968	807	161	20.52	20-22
As in sample 1, Class of 2018 <sup>b</sup>	1089	1078	832	246	20.08	20-22
Reserve Officer Training Core (ROTC) data testing by the Office of Economic and Manpower Analysis <sup>b</sup>	5614	5390	4208	1176	20.01	20-30
Mturk sample of people employed at least half-time; screened for inattentive responding (see Lortie, 2015) <sup>c</sup>	474	459	210 <sup>c</sup>	247 <sup>c</sup>	34.91	19-70
Members of the United States Army enrolled in Officer Candidate School. This sample is a composite of 00 smaller samples tested through the 2013-2015 timespan, also by the Office of Economic and Manpower Analyses <sup>b</sup>	307	276	220	56	25.80	22-90
<b>Overall Total</b>	<b>10618</b>	<b>10318</b>	<b>8049<sup>d</sup></b>	<b>2261<sup>d</sup></b>	<b>21.09</b>	<b>19-90</b>

*Note.*

a. Participants in all samples were required to answer every question to complete the survey.

b. Members of samples 1-4, 5, and 7 were tested by proctors in large groups and took the test on a military survey-administration system (a few hundred cadets who missed the group testing completed it individually). All individuals in the military samples were told that their scores on the test were consequential because the test scores would be used by the Office of Economic and Manpower Analysis to help them choose the best branch of the U.S. Army to enter into once they had graduated (Mayer & Skimmyhorn, in press). Data for the military samples was examined under a data-sharing agreement between the Office of Economic and Manpower Analysis of the United States Army and the University of New Hampshire.

c. Test-takers in the sole non-military sample, Sample 6, completed their tests online through MTurk; they each received small payments, contingent on their completion of all the questions. Participants in sample 6 were permitted to specify a gender other than male or female. Details of this sample can be found in Lortie (2015).

d. The sum of men and women reflects 5 instances for which data was unavailable and 2 participants of sample 6 who self-identified as "other."



**Further screening.** Screening for attentiveness often enhances the quality of survey results and their interpretation (DeSimone, Harms, & DeSimone, 2015). We assigned warning values to the Infrequency and Repetition scales: Participants earned a warning value on the infrequency scale if they endorsed 25% of the least frequently-endorsed alternatives (a random pattern), and one on the repetition scale if they answered with a single letter more than 2/3rds of the time. Roughly 1% and .5% of the sample, respectively, received warnings on the scales. We set caution levels for both scales, assigned to the interval just below the warning level, but still in the extreme tails of the distributions. Parallel cut points (using prorating) were set for the second-half scales. Given the circumstantial evidence that some participants experienced a drop in attention, we screened out those with a warning on either scale, or two cautions on any of the scales; this excluded 300 (2.8%) of the 10,618 participants.

### **Exploratory and Cross-Check Subsamples**

In the factor and item analyses of the scale, we explored and finalized models in an exploratory sample of  $N = 5,144$  odd-numbered participants and then cross-checked them in the remaining  $N = 5,174$  even-numbered participants. (More odd- than even numbered participants were screened out for attentional drop-off). We split the sample odd-even so as to tie the two samples to the original archived data file, to ensure that we represented all samples equally and so the split could be readily re-created as needed.

### **Addressing the Factor Structure of Personal Intelligence (Hypothesis 2)**

We next turned to whether personal intelligence can be adequately represented either as a single overall factor or as two or more highly-correlated factors. To test this hypothesis, we first conducted exploratory and confirmatory factor analyses in the exploratory sample and then checked the final confirmatory model in the cross-check sample.

**Overview of the TOPI 1.4 factor analyses.** For all factor models we evaluated fit with reference to acceptance criteria for the CFI and TLI of “close to” .95 or higher, and for the RMSEA of “close to” .06 or lower (Boomsma, Hoyle, & Panter, 2012, p. 27; Hu & Bentler, 1999). TOPI items are dichotomously scored (i.e., correct or incorrect) and so categorical in form; we therefore used *Mplus*’s weighted least squares, mean and variance adjusted estimation (WLSMV) as it is particularly appropriate for test items of this (Muthén & Muthén, 1998-2015). For our exploratory analyses, we used a facparsim rotation, which is especially suitable for modeling large numbers of items (Finch, 2011; Sass & Schmitt, 2010). Only delta parameterization was available for the exploratory analyses in *Mplus* version 7.2; our confirmatory analyses used WLSMV with theta parameterization because it allowed us to compare modification indices for item pairs (the two parameterizations yielded similar models and fits).

**Initial exploratory factor analyses.** We first compared one, two, and three exploratory factor solutions in the exploratory sample. The fit of each solution to the data is reported in the top rows of Table 2 labeled “Initial Exploratory Factor Analyses.” The one-factor model fit was imperfect with its CFI and TLI closer to .90 than .95. A two-factor solution fit better with CFI = .95; TLI = .95, and RMSEA = .01. An examination of the scree plot also suggested a two-factor solution. The three-factor solution did fit best, but its third factor appeared to be a “bloated specific:” its several items shared wording and came from a single item cluster, rather than representing a more broadly meaningful factor.

Table 2

## Factor Models of the TOPI 1.4: Iterations and Concluding Solutions

Model Tested	Item Split	Variables/ Free parameters	Fit Indices				$r_{factors}$	
			Chi-2	df	RMSEA	CFI		TLI
<b>Initial Exploratory Factor Analyses, Odd Sample (N=5,144)</b>								
<i>1- to 3-Factor Solutions Facparsim-Rotated, Oblique</i>								
One factor model	na	93/93	9813.11	4185	.016	.907	.905	na
Two factor model	43/50	93/185	7298.51	4093	.012	.947	.945	$r_{12} = .46$
Three factor model	34/39/20	93	6488.78	4002	.011	.959	.956	$r_{s_{123}} = .35$ to $.52$
<b>Confirmatory Two Factor Models, Odd Sample (N=5,144)</b>								
<i>Removing Cross-Loading Items (&gt; .25 on both factors)</i>								
Simple structure	80: 40/40	80/161	7127.40	3079	.016	.912	.910	$r_{12} = .79$
<i>Removing Low-Loading Items (&lt; .25)</i>								
Simple structure	72: 36/36	72/145	5354.72	2483	.015	.934	.932	$r_{12} = .80$
<b>Final Factor Model, Odd Sample (N=5,144)</b>								
<i>Removing Items with Poor Pairwise Fit (Large Modification Indices)</i>								
Simple structure	68: 34/34	68/137	4074.63	2209	.013	.952	.950	$r_{12} = .82$
Bifactor Model <sup>a</sup>	68: 68/34/34	68/204	3408.78	2142	.011	.967	.965	$r^{**} = .00$
<i>Configural (Factor) Invariance for Two Factors for Women and Men<sup>b</sup></i>								
Men, separately	68: 34/34	68/137	3632.04	2209	.013	.955	.953	$r_{12} = .82$
Women, separately	68: 34/34	68/137	2454.66	2209	.010	.948	.947	$r_{12} = .82$
Combined model	68: 34/34	69/208	5722.63	4484	.011	.962	.962	$r_{12m/f} = .82; .81$
<b>Confirmatory Two Factor Models, Even Sample Cross-Check (N = 5,174)</b>								
Simple structure	68: 34/34	68/137	4226.71	2209	.013	.947	.945	$r_{12} = .81$
Bifactor Model <sup>a</sup>	68: 68/34/34	68/204	3380.06	2142	.011	.967	.965	$r^{**} = .00^c$
<b>Study 2 Cross Validation Sample (N = 8,459)</b>								
Simple structure	67: 33/34	67/135	5682.48	2143	.014	.957	.956	$r_{12} = .87$

a. The bifactor model employs an overall factor on which all items load, as well as group factors (in this case, two). The overall factor and two specific factors are all constrained to be uncorrelated with one another. In the resulting model, every item is a product of two independent sources of variance: that due to the overall factor (overall personal intelligence, in this instance), and that due to the specific factors.

b. A test of configural invariance specifically indicates whether the number of factors of the model fit for separate groups, in this case, whether it fit equally well for men and women. To carry out the test, the model is first fit to groups of women and men separately, and then fit in a combined model including both groups.

c. As noted in 'a' above, the correlation between the two specific factors is constrained to zero in the bifactor models. The constraint is useful for estimating the variance accounted for by a general factor relative to the specific factors (see text).

**Interpreting the factors.** We next interpreted the factors by examining the items that loaded on them. The first factor included items that asked about: (a) traits that are consistent with one another (e.g., liveliness and extraversion), (b) consistencies between an individual's personal goals and behaviors, and (c) a person's traits and their reactions in a given context, as well as (d) perceiving congruency between a person's behaviors and their motives and traits. We labeled

this first factor Consistency-Congruence (CC) personal intelligence because it concerned reasoning about consistencies among plans, traits and behaviors.

The second factor asked questions about (a) understanding the dynamic interplay between two or more goals and their possible conflicts, (b) knowing how personal memories can motivate (or demotivate) a person, (c) identifying how a given self-perception might promote a behavior, and (d) analyzing alternative perceptions of oneself (or another person), and sensibly integrating the sometimes contradictory information. We labeled this second factor Dynamic-Analytic (DA) personal intelligence because it involved reasoning about personality dynamics and integrating information. Both factors drew items from all four problem-solving areas represented in the test.

**Initial test of the two-factor model.** We next confirmed the two-factor model, constraining each item to load on only a single factor. We first removed 13 items that loaded above  $|.25|$  on both factors and then tested the model with the remaining 80 items, placing the remaining items on the factor on which they had loaded most highly before. The solution fit reasonably well, and the two factors correlated with one another  $r = .79$ , but the CFI and TLI, at around  $.91$ , did not meet our expectation (see Table 2 for details). We next removed all items that failed to load at least  $|.25|$  on their assigned factors, leading to a further slight improvement in fit. Finally, we identified five item pairs that exhibited large modification indices and, on that basis, dropped four items (one item was in two pairs). The two-factor model fit the remaining 68 items with a CFI and TLI of  $.95$ , and an RMSEA of  $.01$ , which met criteria.

We also successfully fit a bifactor model to the same 68 items (see Table 2) and used that to further check of the scale's dimensionality. We calculated the amount of variance of each factor due to the common personal intelligence factor, termed its *explained common variance* or ECV (Reise et al., 2010). The overall personal intelligence factor accounted for 72% of the variance of the first factor and 79% of the second (76% overall). Anything less than 85% is regarded as supporting multidimensionality and therefore supported the use of two separate factors (O'Connor Quinn, 2014; Stout, 1990).

**Configural invariance (Hypothesis 3).** We also checked whether men and women exhibited the two-factor structure in their separate groups; they did, as indicated in the factor invariance row of Table 2.

**Re-confirmation on the cross-check sample.** As a further step, we examined the fit of the model in the cross-check sample ( $N = 5,174$ ). The fit of both the simple structure and bifactor models was almost identical to that found in the exploratory sample (Table 2).

### Item Analyses of the Scales

Having divided the TOPI 1.4 into two related factors, we next conducted item analyses of each scale in IRTPRO (Cai, Thissen, & du Toit, 2016). We employed a 2-parameter IRT model of the scales that estimated each item's difficulty level and power of discrimination in the exploratory sample. In our analyses of the two scales, there was no advantage of a 3-parameter relative to a 2-parameter model, as assessed by the difference in  $-2\log$ likelihoods across models, where  $\chi^2 = -2\ln_{3PL} - 2\ln_{2PL}$ . For the CC scale, these values were  $\chi^2(34) = 141095.85 - 141066.26 = 29.59$ , n.s., and for DA,  $\chi^2(34) = 169466.07 - 169443.66 = 23$ , n.s..

In the results that follow, we corrected significance levels for the number of item comparisons using the Benjamini-Hochberg correction (Williams, Jones, & Tukey, 1999).

**General fit.** For both scales, all items fit the 2-parameter model adequately according to the reasonableness of the parameter estimates and the summed score ( $S-\chi^2$ ) item level

diagnostics. The initial model fit also was indicated by an RMSEA of .02 and .01 for the Consistency-Congruence and Dynamic-Analytic scales.

**Marginal dependence.** For the CC scale, 10 item pairs showed substantial local dependence ( $LD-\chi^2 > 10$ ), and for the DA scale, 13 item pairs, indicating that some item pairs shared variance beyond that of the targeted factor. We examined the content of each of those item pairs. For the CC scale, nine of the pairs addressed reasonably diverse content; the tenth pair, however, employed repetitive wording, and we dropped the weaker of the two items, improving the model fit to an RMSEA of .01. The DA scale pairs exhibited no repetitive content and we retained all item pairs.

**Item difficulty and discrimination across ability levels.** For the two scales, the item slopes ( $a$  parameters) ranged from .48 to 1.48, with a mean of .84 for CC, and from .39 to 1.70, with a mean of .87, respectively for DA. The item difficulty levels ( $b$  parameters) for the CC and DA scales ranged from -3.85 to -.80 and from -3.93 to .07, with a mean of -1.77 respectively, indicating that most test items could be passed by test-takers who are average or above in personal intelligence.

**DIF Analysis for Men and Women (Hypothesis 3 redux, applied to the scales).** Women scored approximately .30 standard deviations above men on the CC scale and .22 higher than men on the DA scale overall. Analyses exploring differential item function (DIF) indicated one item was significantly harder for women than for men on the CC scale; we retained the item because it did not appear to inquire about anything inherently gender-related. On the DA scale, five items were harder for women, and another five harder for men. For the most part the gender differences emerged among men and women with comparatively low ability-levels in the area, and were possibly linked to the gender of the protagonist of the item (gender of the person referred to varied across questions); one item difference might have been due to differential anger responses for women and men. As the differences in the 10 flagged items cancelled one another out, we made no alterations to the scale on this basis.

**Confirmation of findings across samples.** The scales exhibited very similar characteristics in the cross-check sample compared to the exploratory: for example, the average  $a$  parameter of the CC scale differed across samples  $M_{diff} = .02$ , the  $b$  parameter  $M_{diff} = -.02$ .

### Characteristics of the Final Scales

**Scaled scores.** We created scaled scores using a T-scale for both the Consistency-Congruency and Dynamic-Analytic scales based on the cross-check sample; that scaling set the  $M = 50$  and  $S = 10$ .

**Test and Scale Correlations.** The top of Table 3 shows the correlation between the Consistency-Congruence and Dynamic-Analytic scales is  $r = .58$  for the cross-check sample. (The results for all participants are in the second portion of Table 3). A composite score that averages the two scaled scores (CA and DA), referred to as the TOPI 1.4R Total also is reported, along with the original TOPI 1.4 total (all 93 items).

**Better discrimination at low levels of performance (Hypothesis 4).** Consistent with the idea that the scales better discriminated among test-takers at lower levels of performance, distributions of test-takers were negatively skewed for both scales (cross-check sample: CC = -1.51; DA = -1.26,  $ps < .001$ ) and standard errors of measurement for the scaled scores were smaller for intervals below the mean than above it for both T-scales (cross-check sample for CC:  $SEM_{below} = 4.5$ ;  $SEM_{above} = 6.3$ ; for DA:  $SEM_{below} = 4.3$ ;  $SEM_{above} = 5.9$ ).

**Scale reliabilities (Hypothesis 5).** As implied by all the analyses so far, the two IRT-based scales and total (averaged) scale all exhibited reasonable reliability under models of both

classical test theory and item response theory. Coefficient alpha reliabilities in ranged from .74 to .85 in the cross-check sample. The IRT-based marginal reliabilities were somewhat lower at .66 and .74 because of the weaker discrimination of the scales among higher-ability test-takers (see Table 3). The middle portion of Table 3 reports the means and standard deviations of the scaled scores and their composite for Study 1.

Table 3

Means, Reliabilities, and Correlations for the TOPI 14R *Consistency-Congruence*, *Dynamic-Analytic* and *Composite* scales

TOPI 14R Scales	Descriptive Statistics			Reliabilities		Correlations		
	Total Mean (S)	Men <sup>1</sup> Mean (S)	Women <sup>1</sup> Mean (S)	Alpha <sup>3</sup>	Margin. (IRT) <sup>5</sup>	Consis.	Dyn.	Orig. TOPI 14
Study 1 Cross-Check (Even-Numbered) Sample, <i>N</i> = 5,174								
Consistency	49.69 (8.08)	49.37 (8.19)	50.85 (7.57)	.74	.66	1.00**		
Dynamic	49.83 (8.53)	49.56 (8.53)	50.76 (8.49)	.80	.74	.58**	1.00**	
Composite <sup>2</sup>	49.76 (7.38)	49.46 (7.45)	50.80 (8.49)	.85** <sup>4</sup>	Not est. <sup>6</sup>	.88**	.90**	.96**
Study 1 Total Sample, <i>N</i> = 10,318								
Consistency	49.85 (8.12)	49.47 (8.21)	51.19 (7.64)	.75	.66	1.00**		
Dynamic	49.92 (8.58)	49.59 (8.61)	51.06 (8.36)	.80	.74	.59**	1.00**	
Composite <sup>2</sup>	49.88 (7.45)	49.53 (7.52)	51.13 (7.06)	.85** <sup>4</sup>	Not est. <sup>6</sup>	.89**	.90**	.97**
Study 2 Total Sample, <i>N</i> = 8,459								
Consistency	49.47 (8.65)	49.12 (8.78)	50.69 (8.08)	.79	.65	1.00**		
Dynamic	49.24 (9.01)	48.93 (9.07)	50.30 (8.76)	.82	.74	.64**	1.00**	
Composite <sup>2</sup>	49.36 (7.99)	49.02 (8.10)	50.49 (7.48)	.88** <sup>4</sup>	Not est. <sup>6</sup>	.90**	.90**	.97**

<sup>1</sup>Men and women number 6843 and 1971 in the cross-check sample, 8049 and 2261 in the Study 1 total sample, and 6539 and 1920 in the Study 2 Replication Sample. Men and women do not add to the total in Study 1 because in one subsample, test-takers could endorse an “other” alternative.

<sup>2</sup>Formed from the mean of the scaled scores of the Consistency-Congruency and Dynamic-Analytic factors (see text).

<sup>3</sup>Based on standardized items.

<sup>4</sup>This estimate is the coefficient alpha on the standardized items of the composite unscaled 67 items; the alpha for the combined scaled scores may vary slightly from this value but cannot be directly calculated.

<sup>5</sup>The marginal reliabilities are for the scaled scores on the summed scores (SS/SS). The estimates based on the response pattern scoring (RPS) were trivially higher.

<sup>6</sup>The marginal reliabilities of the simple summed score of two IRT-based scaled scores cannot be estimated.

### Correlations between Actual Personal Intelligence and Self-Estimated Personal Intelligence (Hypothesis 6)

We further had hypothesized that the *Test of Personal Intelligence* measures mental ability by drawing on cognitive abilities, distinct from self-judgment. The master data file also included scores on the Self-Estimated Personal Intelligence test (SEPI) for 8,866 cases. The 16 item scale exhibited an alpha reliability of  $\alpha = .87$  and test-to-test correlations with Consistency-Congruence and Dynamic-Analytic abilities at levels of  $r = .24$  and  $.24$ , with the TOPI Total at  $r = .26$ , and the original 93 item TOPI 1.4,  $r = .28$ ,  $ps < .01$ , levels that fell slightly higher than our

predicted levels. We will revisit the TOPI scales' validities in Study 3 and the General Discussion.

### **Study 2: Cross-Validation in an Independent Sample**

In Study 2 we retested the two-factor model and reexamined the scales in an independent, archival file of  $N = 8,814$ . All data were from the *Office of Economic and Manpower Analysis*, and delivered between November 2015 to September 2016. These latter test responses were scored for the OEMA but otherwise unexamined until late September of 2016, well after the analyses for Study 1 were complete.

#### **Participants**

Participants in Study 2 were  $N = 8,814$  test-takers from two further classes from West Point ( $Ns = 973$  and  $1,107$ ) and two further samples of ROTC students from different years ( $Ns = 5,512$  and  $1,222$ ), the latter including only ROTC scholarship students. There were 6,843 men and 1,971 women with a mean age of 20.07 and a similar ethnic distribution as before.

#### **Measures, Procedures, and Screening**

Our measures, procedures, and screening were the same as in Study 1.

#### **Results**

**Screening.** In Study 2, the screening (unchanged from Study 1) resulted in the removal of 355 test-takers or 4%, a rate similar to but somewhat higher than the 2.8% exclusion rate in Study 1, leaving  $N = 8,459$ .

**Test of the Two-Factor Model.** A test of the two factor simple-structure model on the final 67 item scale indicated a similar fit as before of  $CFI = .96$ ;  $TLI = .96$ , and  $RMSEA = .014$ . In Study 2; the correlation between the two factors rose to  $r = .87$  (from  $.82$  in Study 1) due to a slightly greater range of test-taking ability among participants: If we correct the correlation in the Study 1 sample for restriction of range relative to Study 2, the  $r = .86$  between factors is almost equivalent. The Explained Common Variance continued to indicate the presence of two factors, at 83% and 81%, as tested within the context of a bifactor model. In our reexamination of the item response analyses, the original model fit well ( $RMSEAs = .02$  and  $.02$ ) and no modifications to the two scales were regarded as necessary.

### **Study 3. Examining the Correlation of the TOPI 1.4R Scales with Selected Criteria: Reanalysis of Mayer, Panter & Caruso, 2012 (Study 3)**

To understand more about the *Consistency-Congruence* and *Dynamic-Analytic* scales developed here (and their composite), we rescored the subset of 67 items of the TOPI 1.2 used in Mayer et al. (2012), Study 3, that now form the Consistency-Congruency and Dynamic-Analytic scales of the TOPI 1.4R. The earlier Study 3 had been conducted to correlate the TOPI with other psychological tests measuring intellectual ability, the big five, and other qualities described below. Our two hypotheses in this reanalysis of earlier data were that (a) the two newly formed subscales would exhibit reliabilities and correlate with one another approximately as in Studies 1 and 2 and (b) that the two scales would exhibit differential correlations with at least several criterion measures used in the original Study 3.

#### **Method and Procedure**

We reanalyzed the data involving  $N = 384$  ethnically diverse college students (52.8%

women, 47.2% men) who had completed a vocabulary test, measures of the Big Five, a scale of psychological mindedness, and a few measures described below.

The scale items of the TOPI 1.4R all had been included in the TOPI 1.2, and so both scales were completely represented in this reanalysis.

## Results

**Scale Characteristics.** The overall reliabilities for the TOPI 1.4R *Consistency-Congruence* and *Dynamic-Analytic* scales were  $r = .79$  and  $.82$ , and  $.88$  for the whole test. The TOPI 1.4R composite correlated  $r = .97$  with the complete TOPI 1.2 (Table 3, bottom rows).

**Criterion Correlations.** Table 4 also includes the correlations between the two subscales and criteria from the earlier study. Both scales and the composite correlate significantly with a number of the criterion measures at roughly the same levels as had the TOPI 1.2 in the original report. The TOPI 1.4R scales correlate positively with related mental abilities in the  $r = .37$  to  $.64$  range,  $ps < .001$ . They correlate with Agreeableness and Conscientiousness in the Big Five  $r = .12$  to  $.20$ , with Psychological Mindedness  $r = .29$  and  $.39$ , and negatively with symptoms of personality disorders  $r = -.07$  to  $-.22$ . They show signs of correlating with lifespace data (i.e., questions related to specific behavioral interactions and decisions), such that people low in personal intelligence exhibited more judgmental reactions to people and less future planning.

Among these criteria—which were selected to correlate with overall PI and not to distinguish between abilities—the Consistency-Congruency and Dynamic-Analytic scales mostly performed similarly, although the DA scale correlated more highly with other mental abilities than did CC, as indicated in the block of ability measures at the top. Beyond that, there is just a hint that CC may be less protective against the manifestations of narcissistic grandiosity given its lower (negative) correlation with the scale of that name, and its lower relation to overconfidence in judging people, as illustrated by the lifespace item “...discover you were right” about turning someone down as a roommate. We further consider the two scales and their possible differences in the General Discussion.

## General Discussion

### Summary of Findings

Earlier research has supported a view of personal intelligence as a measure of a person-centered intelligence that correlates with other scales and with performance in school and on-the-job (Mayer & Skimmyhorn, in press; Mayer et al., 2012). Here, facilitated by a now larger sample of test-takers, we examined the ability in more detail than has been done in the past: We explored whether there existed subsidiary mental skills that made up personal intelligence, its relation to self-judged insight into personality, and the distribution of the ability.

We found evidence that people can be characterized as employing two closely-related abilities to answer questions in the area, Consistency-Congruence and Dynamic-Analytic reasoning, and that these were largely unrelated to how smart a test-taker thought she was about people, as measured by a self-judgment test of personal intelligence. Other findings included that the TOPI 1.4R scales better distinguish among test takers at low rather than high ability levels, and that it made sense to add several scales to the TOPI 1.4R to monitor attentive responding.

Table 4

The TOPI 1.4R and Criteria: Reanalysis of Data from Mayer, Caruso & Panter, 2012 ( $N = 384$ )

TOPI and Criterion Measures	TOPI 14R Factor Scales <sup>a</sup>		Difference (absolute) <sup>a,b</sup>	Composite Score <sup>a</sup>
	Consistency	Dynamic		
<b>TOPI Scales and Other Mental Abilities</b>				
TOPI-Consistency-Congruence	1.00			
TOPI-Adaptive-Dynamic	.67**	1.00		
TOPI 1.4-R Composite	.89**	.93**	.04**	1.00
Vocabulary	.31**	.44**	.13**	.42**
Reading the Mind in the Eyes	.37**	.50**	.13**	.49**
MSCEIT-Strategic Area	.56**	.63**	.07*	.66**
Understanding Task	.37**	.51**	.14**	.48**
Managing Task	.54**	.64**	.10**	.65**
<b>Big Five</b>				
Extraversion	-.01	-.06	.05	-.04
Neuroticism	-.06	-.03	.03	-.05
Openness	.06	.10	.04	.09*
Agreeableness	.12*	.15**	.03	.15**
Conscientiousness	.19**	.16**	.03	.20**
<b>Psych Minded-Total</b>				
Discussing prob.	.29**	.39**	.10**	.38**
Access feelings	.25**	.33**	.08*	.33**
Understanding	.18**	.21**	.03	.22**
Motivation	.12*	.15**	.03	.15**
Open Change	.24**	.29**	.05	.29**
	.09	.15**	.06	.13*
<b>Symptomatology</b>				
Maladaptive Agreeableness	-.14**	-.15**	.01	-.16**
Narcissistic Grandiosity	-.13*	-.22**	.09*	-.19**
Narcissistic Personality (NPI)	-.07	-.13*	.08	-.11**
<b>Self-Described Social Skills</b>				
Initiating relationships	.02	-.02	.04	-.06
Providing emotional support	.16**	.14**	.02	.16**
Asserting influence	.09	.02	.07	.06
Self-disclosure	-.00	-.01	.01	-.01
Conflict resolution	.08	.07	.01	.08
<b>Life Space Questions (“Over the past week...did you:”)</b>				
a. Watch yourself do something to improve?	-.16**	-.19**	.03	-.19**
b. Ask for feedback?	-.11*	-.10	.01	-.12*
c. Read about role model?	-.20**	-.23**	.03	-.24**
d. Plan for your future?	.12*	.15**	.03	.15**
e. Turn down someone to be a roommate and discover you were right?	-.27**	-.36**	.09*	-.35**
f. Describe someone’s personality in detail in an e-mail?	-.20**	-.22**	.02	-.23**

<sup>a</sup>\* $p < .05$ , \*\* $p < .01$ ; <sup>b</sup>significance levels are a consequence of both difference and correlation level



**The Two-Factor Model of the Test—And a One-Factor Representation as an Alternative Consistency-Congruence and Dynamic-Analytic Reasoning.**

The two factor model divides people's personal intelligence into Consistency-Congruence and Dynamic-Analytic personal intelligence. These factors showed up in exploratory factor analyses of the scale, and could be modeled well in confirmatory analyses using a simple correlated factor structure. People use their Consistency-Congruence reasoning to think about traits, how traits relate to one another, and how they predict people's actions. People use their Dynamic-Analytic reasoning to understand how different parts of personality work together, how sometimes contradictory information about personality can be integrated, how different people can perceive one another differently, and how to set goals for the future.

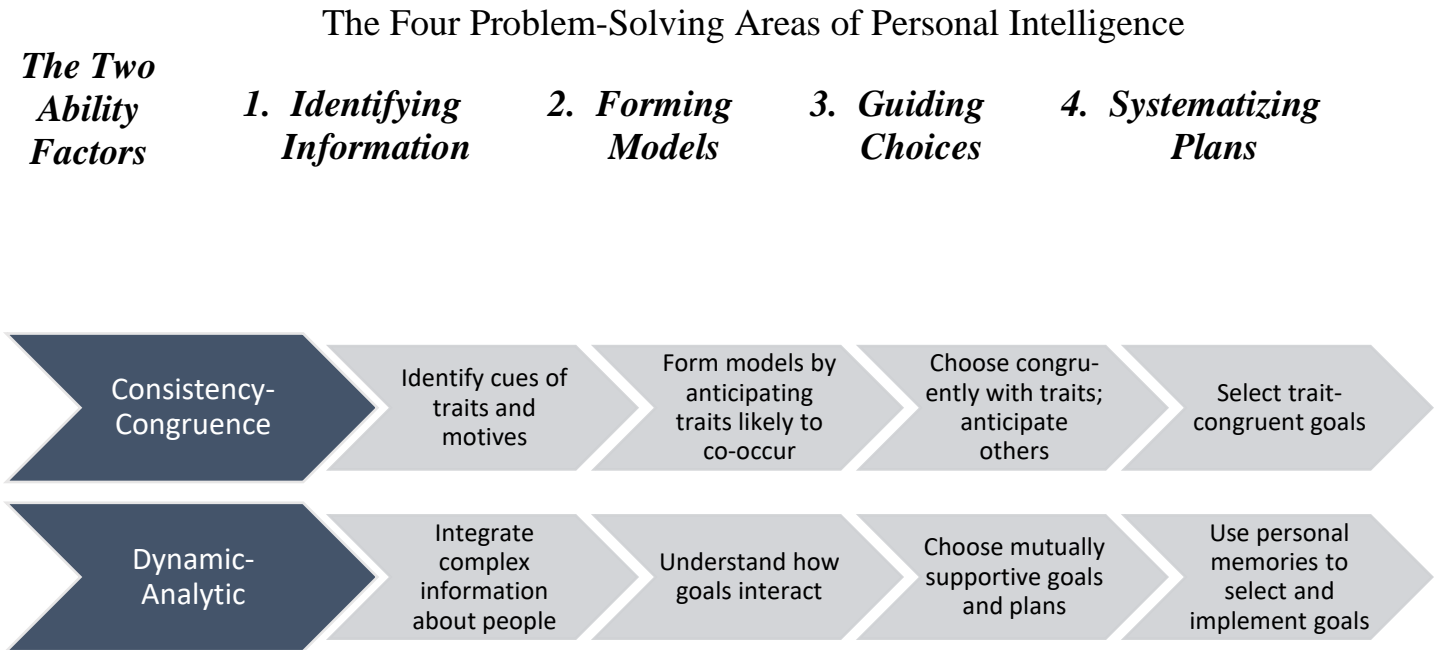
**A One Factor Alternative.** Although we like the two-factor approach, we freely acknowledge that alternative models also fit the data well, especially the legitimacy of a "one intelligence" approach to the personal intelligence area, as supported by the bifactor model, and that other researchers may prefer it (Reise et al., 2010, p. 554). For that reason, we also have calculated a composite scale that takes the average of the Consistency-Congruence and Dynamic-Analytic scaled scores, weighing them evenly in the result (to equally represent the abilities in the area, as now understood). This TOPI Total Scale may be convenient for use when a simple summary variable is needed to represent personal intelligence.

**The correlation between Consistency-Congruence and Dynamic-Analytic personal intelligence.** Indeed, a complication of the two-factor approach is that the CC and DA scales possess an estimated correlation of between  $r = .81$  and  $.82$  in Study 1 to  $.87$  in Study 2 if measured with perfect weighting of parameters and perfect reliability. The actual obtained correlations between the two scales in our studies ranged from  $r = .58$  to  $.67$ . Although the estimated values seem high, they fall squarely within the range of similar estimates found for correlations among other broad intelligences: MacCann et al. (2014, Table 5) found estimated correlations between crystallized and fluid intelligences of  $r = .87$ , between verbal and fluid intelligence of  $r = .88$ , and between verbal intelligence and quantitative reasoning of  $r = .75$ ; Kranzler & Keith (1999) estimated correlations among broad abilities measured by the Cognitive Assessment System from  $r = .61$  to  $.93$ .

**The Relation between the Four Problem-Solving Areas of Personal Intelligence and the Two-Factor Solution.**

Earlier we presented a model of four problem solving areas of personal intelligence: identifying information, modeling personality, guiding choices and systematizing goals. We regarded those problem-solving areas as likely distinct from the abilities people apply to the problem areas: Educators rightly distinguish between fiction and nonfiction when teaching literature, but their students use vocabulary skills and sentence comprehension to understand both. The situation for personal intelligence is depicted in Figure 1. There, the four problem-solving areas are indicated across the top, and the two reasoning areas are depicted in arrow-like figures that span the problem areas. Consistency-Congruence reasoning, for example, can be used to identify trait-related clues in the Identifying Information area, or to select objectives consistent with one's traits in the Systematizing Goals area. By comparison, Dynamic-Analytic reasoning can be used to integrate contradictory information about a person's traits in the Identifying Information area, or to ensure one chooses goals that are mutually supportive where possible in the realm of Systematizing Plans.

Figure 1: Consistency and Dynamic Personal Intelligences Operate across the Four Problem-Solving Areas of Personal Intelligence



**Differential Predictions from the Two Scales.**

The differential prediction of the two factors is as-of-yet mostly unexplored, and limited by their approximately  $r = .85$  correlation with one another. That said, Dynamic-Analytic reasoning appears more closely related to other intelligences and Consistency-Congruence may be more related to other qualities we might speculate about: One possibility is that CC reasoning correlates with people’s implicit personality theories about how unchangeable or adaptable people are. Carol Dweck and colleagues (Olson & Dweck, 2009; Plaks, Levy, & Dweck, 2009) have characterized a portion of the population as applying a *fixed* mindset to how people behave, emphasizing the stable nature of people’s personality; this may relate to abilities assessed by the Consistency-Congruence scale. By comparison, others employ a *growth*-oriented mindset emphasize the changing nature of personality, which may in turn be more related to the Dynamic-Analytic scale.

**The Fifth Person in the Room (and the Distribution of Personal Intelligence)**

More people score highly on the TOPI than low—its scores are negatively skewed—and, like other tests that examine reasoning about people, it also discriminates more clearly among people low in ability than those who are high (e.g., Maul, 2012). It is worth considering this distribution of abilities on the test. To use the Dynamic-Analytic scale as an example, the top fifth of test takers get between 90%-100% (31 to 34 items) of the questions correct; if this were a classroom, they would be the “A” students. The next 30% of test-takers, all still above average, correctly answer 79% to 89% of the time (27 to 34 items)—they would be the “B” students. The

following 30%—the roughly third of the sample somewhat below-average, answer correctly 68% to 78% of the time (23 to 26 items correct)—most of this group are the “C” students—and we could fairly say that although they perform all right, they “don’t seem to get” a good deal about themselves and others. Then there is the bottom 20%, who scored between 12% to 65% correct (from 4 to about 22 items). The lowest among this group scored below chance levels, perhaps from bad luck, and the rest simply couldn’t answer anything (fewer than half a percent were at this extreme). Note that these lowest-scoring individuals were apparently still paying attention: They avoided the relatively improbable alternatives that are flagged by the Infrequency Index (part of the Inattentiveness Warning Scale). The rest of this group struggled between commonly-endorsed but incorrect answers and the correct answers, performing between 25% to about 65% correct.

Understanding this distribution may afford us some insight into our daily experiences with other people. The results indicate that, among a group of any 10 people there will be (on average) one who excels at understanding others, 2 more who are very good at it, and 2 others who generally “get” other people. Among the 5 remaining people are 3 who are rather slow to pick up on the nature of themselves and other people, and two who recognize very little, perhaps nothing, about personality and differences in people at all. Because every fifth person scores fewer than 65% correct of questions about personality, we might refer to them as the “fifth person in the room.” Their test performance suggests that they are fairly often confused as to people’s individual traits and dynamics, and may, for example, mistake shyness for aloofness, attribute deceit to innocent forgetfulness, or pointlessly continue to criticize people who react to them defensively. They also may be unaware of their deficit: Self-estimated personal intelligence exhibited minimal relations with actual ability in Studies 1 and 2, calling to mind, from Shakespeare’s *Measure for Measure*, Isabella’s remark that a judge might be “Most ignorant of what he’s most assur’d” (*Measure for Measure*, Act II, Scene 2, line 117). It is possible that a person with a lifetime of misunderstanding of personality could feel confusion and frustration, which could account for why some low-scoring participants exhibited relatively higher symptoms of personality disorders.

### **Educational Implications**

One purpose of identifying subsidiary abilities in the personal intelligence areas was to better understand how we might help people to increase their knowledge and effective performance in these areas. The presence of Consistency-Congruency reasoning argues for teaching people about traits, their meanings and variations, as well as their relationships to behavioral outcomes. By comparison, the presence of a Dynamic-Analytic ability suggests that people might be taught about how different parts of personality affect one another: When a person’s goals are consistent or inconsistent, or when motives and goals diverge. We believe that most people have sufficient ability in the area to benefit from curricular-based education and training, and to become better at such problem-solving should they choose to.

### **Future Directions**

There remains much to learn about personal intelligence. From a practical standpoint, it may be possible to develop future versions of the *Test of Personal Intelligence* that measure more precisely at the higher end of the ability continuum. In addition, people engage in many kinds of problem solving about one another that may be added to the TOPI or tests like it. As new kinds of test items are added, new subsidiary abilities in the area may be uncovered. The

TOPI represents an advance, however, in that it serves to assess personal intelligence and to clarify a test takers's strengths and weaknesses in the fundamental abilities it measures.

We believe that the *Test of Personal Intelligence* is, in its present form, sufficiently well worked out to promote continued exploration regarding the real-life criteria that personal intelligence may predict. We believe that it must be very different to be among the roughly half of all test takers who appear to understand personality well versus the lower half of test takers who appear relatively confused and misdirected when understanding personalities—be it their own or another person's. Surely these differences in understanding oneself and others are likely to affect a person's life course over time.

### References

- Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The 'reading the mind in the eyes' test revised version: A study with normal adults, and adults with asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, 42(2), 241-251. doi:10.1111/1469-7610.00715
- Boomsma, A., Hoyle, R. H., & Panter, A. T. (2012). The structural equation modeling research report. In R. H. Hoyle, & R. H. (. Hoyle (Eds.), (pp. 341-358). New York, NY, US: Guilford Press.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, 111(4), 1061-1071. doi:10.1037/0033-295X.111.4.1061
- Brackett, M. A., & Mayer, J. D. (2003). Convergent, discriminant, and incremental validity of competing measures of emotional intelligence. *Personality and Social Psychology Bulletin*, 29(9), 1147-1158. doi:10.1177/0146167203254596
- Buss, D. M. (2008). Human nature and individual differences: Evolution of human personality. In L. A. Pervin (Ed.), *Handbook of personality psychology: Theory and research (3rd ed.)*. (pp. 29-60). New York, NY US: Guilford Press.
- Cai, L., Thissen, D., & du Toit, S. H. C. (2016). *IRTPRO : Flexible, multidimensional, multicategorical IRT modeling (version 3)*. Lincolnwood, IL: Scientific Software International.
- Colom, R., Lluís-Font, J., & Andrés-Pueyo, A. (2005). The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutrition hypothesis. *Intelligence*, 33(1), 83-91. doi:10.1016/j.intell.2004.07.010
- DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2015). Best practice recommendations for data screening. *Journal of Organizational Behavior*, 36(2), 171-181. doi:10.1002/job.1962
- DeYoung, C. G. (2015). Cybernetic big five theory. *Journal of Research in Personality*, 56, 33-58. doi:10.1016/j.jrp.2014.07.004
- Dunbar, R. I. M. (2009). The social brain hypothesis and its implications for social evolution. *Annals of Human Biology*, 36(5), 562-572. doi:10.1080/03014460902960289
- Finch, W. H. (2011). A comparison of factor rotation methods for dichotomous data. *Journal of Modern Applied Statistical Methods*, 10(2), 549-570.
- Fiori, M., Antonietti, J., Mikolajczak, M., Luminet, O., Hansenne, M., & Rossier, J. (2014). What is the ability emotional intelligence test (MSCEIT) good for? an evaluation using item response theory. *PLoS ONE*, 9(6), 1-11. doi:10.1371/journal.pone.0098827
- Hood, S. B. (2009). Validity in psychological testing and scientific realism. *Theory & Psychology*, 19(4), 451-473. doi:10.1177/0959354309336320

- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1-55. doi:10.1080/10705519909540118
- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99-114. doi:10.1007/s10869-011-9231-8
- Joint\_Committee. (2014). *Standards for educational and psychological testing*. Washington, DC US: American Psychological Association.
- Kranzler, J. H., & Keith, T. Z. (1999). Independent confirmatory factor analysis of the cognitive assessment system (CAS): What does the CAS measure? *School Psychology Review*, 28(1), 117-144.
- Larsen, R. J., & Buss, D. M. (2014). *Personality psychology: Domains of knowledge about human nature*. Boston, MA: McGraw Hill.
- Lortie, B. (2015). *Personal intelligence at work*. (Unpublished Senior honors thesis). University of New Hampshire, Durham, NH.
- MacCann, C., Joseph, D. L., Newman, D. A., & Roberts, R. D. (2014). *Emotional intelligence is a second-stratum factor of intelligence: Evidence from hierarchical and bifactor models*. US: American Psychological Association.
- Maul, A. (2012). The validity of the Mayer–Salovey–Caruso emotional intelligence test (MSCEIT) as a measure of emotional intelligence. *Emotion Review*, 4(4), 394-402. doi:10.1177/1754073912445811
- Mayer, J. D., & Skimmyhorn, W. (in press). Broad intelligences and the big five predict cadet performance at west point. *Journal of Research in Personality*, 00, 000-000.
- Mayer, J. D. (2008). Personal intelligence. *Imagination, Cognition and Personality*, 27(3), 209-232. doi:10.2190/IC.27.3.b
- Mayer, J. D. (2014). *Personal intelligence: The power of personality and how it shapes our lives*. New York: Scientific American / Farrar Strauss & Giroux.
- Mayer, J. D., Caruso, D. R., & Salovey, P. (2016). The ability model of emotional intelligence: Principles and updates. *Emotion Review*, 8, 1-11.
- Mayer, J. D., Panter, A. T., & Caruso, D. R. (2012). Does personal intelligence exist? evidence from a new ability-based measure. *Journal of Personality Assessment*, 94, 124-140. doi:10.1080/00223891.2011.646108
- Mayer, J. D., Panter, A. T., & Caruso, D. R. (2014). *Test of personal intelligence (TOPI 1.4) manual*. (). Durham, NH: University of New Hampshire.
- Mayer, J. D., Panter, A. T., & Caruso, D. R. (2016). *Technical supplement to "A closer look at personal intelligence"*. Unpublished manuscript.
- Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT) users manual*. Toronto, Ontario: Multi-Health Systems.
- O'Connor Quinn, H. (2014). *Bifactor models, explained common variance (ECV) and the usefulness of scores from unidimensional item response theory analyses*. (Unpublished Ph.D.). University of North Carolina, Chapel Hill, NC.
- Olson, K. R., & Dweck, C. S. (2009). Social cognitive development: A new look. *Child Development Perspectives*, 3(1), 60-65. doi:10.1111/j.1750-8606.2008.00078.x
- O'Sullivan, M. (2007). Trolling for trout, trawling for tuna: The methodological morass in measuring emotional intelligence. In G. Matthews, M. Zeidner, R. D. Roberts, G. Matthews, M. Zeidner & R. D. Roberts (Eds.), *The science of emotional intelligence: Knowns and*

- unknowns*. (pp. 258-287). New York, NY US: Oxford University Press. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2007-12449-010&site=ehost-live>
- Paulhus, D. L., Lysy, D. C., & Yik, M. S. M. (1998). Self-report measures of intelligence: Are they useful as proxy IQ tests? *Journal of Personality*, *66*(4), 525-554. doi:10.1111/1467-6494.00023
- Plaks, J. E., Levy, S. R., & Dweck, C. S. (2009). Lay theories of personality: Cornerstones of meaning in social cognition. *Social and Personality Psychology Compass*, *3*(6), 1069-1081. doi:10.1111/j.1751-9004.2009.00222.x
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667-696. doi:10.1080/00273171.2012.715555
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544-559. doi:10.1080/00223891.2010.496477
- Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research*, *45*(1), 73-103. doi:10.1080/00273170903504810
- Sinharay, S., Puhon, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice*, *30*(3), 29-40. doi:10.1111/j.1745-3992.2011.00208.x
- Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, *55*(2), 293-325. doi:10.1007/BF02295289
- Von Bertalanffy, L. (1951). Theoretical models in biology and psychology. *Journal of Personality*, *20*, 24-38. doi:10.1111/j.1467-6494.1951.tb01511.x
- Williams, V. S. L., Jones, L. V., & Tukey, J. W. (1999). Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of Educational and Behavioral Statistics*, *24*(1), 42-69. doi:10.2307/1165261