Personality and Individual Differences 111 (2017) 301–311

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

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

John D. Mayer [a,*], A.T. Panter [b,**], David R. Caruso [c,***]

[a] Department of Psychology, University of New Hampshire, United States
[b] L.L. Thurstone Psychometric Laboratory, Department of Psychology and Neuroscience, The University of North Carolina at Chapel Hill, United States
[c] Yale College Dean's Office, Yale University, United States

ARTICLE INFO

Article history:
Received 11 November 2016
Received in revised form 25 January 2017
Accepted 2 February 2017
Available online xxxx

Keywords:
Personal intelligence
Intelligence
Personality
Test of Personal Intelligence

ABSTRACT

Personal intelligence involves the capacity to reason about personality and personality-related information. Studying ability-based measures of personal intelligence creates a virtuous cycle of better measurement and better theoretical understanding. 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 a Consistency-Congruency factor that concerned understanding traits and their associated behaviors, and a Dynamic-Analytic factor that involved understanding personality processes and goals. The finding cross-validated in Study 2 ($N = 8{,}459$). In Study 3 ($N = 384$), we examined correlates of the two factors. Understanding the abilities involved in personal intelligence may help us to educate people about how to better solve problems about personality.

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). Personality organizes an individual's motives and emotions, knowledge and intelligences, and awareness and self-control (DeYoung, 2015; Larsen & Buss, 2014). Individuals then express their inner personalities in the outer world, through their choices and behaviors (Mayer, 2015; but see Hogan & Foster, 2016, for an alternative view).

People vary in the degree to which they comprehend personality. Psychologists believe that, among our evolutionary ancestors, individuals who better understood themselves and the people around them experienced adaptive advantages relative to others in terms of both survival (e.g., selecting better hunting partners) and reproduction (e.g., better mates, Buss, 2008; Dunbar, 2009). Consistent with that idea, psychotherapists have proposed that certain among their clients possessed *psychological mindedness*—a higher aptitude relative to other people for learning about themselves and others (Appelbaum, 1973). Gardner (1983) described intra- and interpersonal intelligences that included skills for building a coherent identity and understanding other people. And Funder (2001) argued for the existence of a *good judge*, who could perceive the personality characteristics of other people more accurately than average. Such concepts share a common focus on the capacity to reason about personality and personality-related information. Mayer (2008, 2014) suggested that a *personal intelligence* (the term was parallel to social and emotional intelligences) might describe this core ability.

People use their personal intelligence, according to the theory, to (a) identify personality-relevant information in themselves and others and to "read" people's traits; (b) to form models of personalities so as to understand themselves and others, (c) to guide their own and others' choices by setting goals consistent with their interests and values, and (d) to systematize their plans so as to achieve their aims (Mayer, 2008).

To test whether personal intelligence could be objectively measured, the present authors developed a *Test of Personal Intelligence*, or *TOPI*, consisting of questions corresponding to those four areas of problem solving just described, and that yielded scores keyed to each of those areas. Our work with the TOPI indicated that ability-based items about personality could be written, correct answers identified, and that people exhibited reliable individual differences in the reasoning capacities assessed (Mayer, Panter, & Caruso, 2012).

This personal intelligence appears to be a mental ability midway in breadth between *general intelligence* (i.e., general abstract reasoning) and specific abilities to solve narrower problems. Intelligence researchers often describe a continuum of intelligences from the most

☆ This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
* Correspondence to: J.D. Mayer, Department of Psychology, McConnell Hall, 15 Academic Way, University of New Hampshire, Durham, NH 03824, United States.
** Correspondence to: A.T. Panter, L. L. Thurstone Psychometric Laboratory and the Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, NC 27599, United States.
*** Correspondence to: D.R. Caruso, Yale College Dean's Office, Yale University, New Haven, CT 06520, United States.
E-mail addresses: jack.mayer@unh.edu (J.D. Mayer), panter@unh.edu (A.T. Panter), david.caruso@yale.edu (D.R. Caruso).

http://dx.doi.org/10.1016/j.paid.2017.02.008
0191-8869/

general to the specific. The mid-level intelligences, which include verbal, spatial, and quantitative intelligences (among others) are called *broad intelligences* in these models (Carroll, 1993; MacCann, Joseph, Newman, & Roberts, 2014; McGrew, 2009; but see Michell, 2012 for an alternate view).

Certain broad intelligences involve reasoning about things, such as quantitative intelligence, that concerns the manipulation of numbers, or spatial intelligence, that concerns the rotation of objects in space; others such as emotional intelligence (assessed as a mental ability) are more people-focused (Mayer & Skimmyhorn, 2017). Personal intelligence is likely among these people-centered intelligences, concerned as it is with traits of social expression, self-control, and mental abilities, as opposed to more impersonal topics. The TOPI total score correlates with other broad intelligences such as the verbal, quantitative and spatial in the $r = .17$ to .30 range (Mayer & Skimmyhorn, 2017), but its relation with people-centered intelligences is stronger: $r = .53$ with “Reading the Mind in the Eyes,” and $r = .68$ with the Understanding Emotions section of the Mayer-Salovey-Caruso Emotional Intelligence test (Mayer et al., 2012). The degree to which emotional and personal intelligences are psychometrically distinct requires further exploration; that said, the two are conceptually distinct in that it is possible to write hundreds of test items about personality without reference to emotion understanding. Recent findings indicate that personal intelligence predicts consequential academic and performance outcomes incrementally above measures of general intelligence alone, suggesting its practical usefulness in high-stakes testing (Mayer & Skimmyhorn, 2017).

1. The Measurement of Personal Intelligence and Aims of the Present Research

The TOPI employs four scales keyed to the four problem-solving areas of personal intelligence: (a) identifying personality-relevant information, (b) forming models of personality, (c) guiding choices, and (d) systematizing plans. The four test scales, however, overlapped in content more than was optimal for separate indicators of the construct. A first confirmatory factor analysis indicated the scales were difficult to distinguish from one another. The initial factor model was an “imperfect, first representation of the results”—a promissory note that required further consideration (Mayer et al., 2012, p. 136).

We fulfill that promissory note in this article in the form of an item-level factor analyses of the TOPI—an "item"-ized payment—to investigate whether there are alternative, useful approaches to representing the test content. If interpretable factors appear independent of the four content areas, it may suggest new ways of conceptualizing the kind of reasoning that make up the ability. As an analogy to verbal intelligence, people who study literature often make a useful distinction between reading fiction and nonfiction. At the same time, however, people's ability to understand language can be represented according to vocabulary skill and sentence comprehension. In other words, experts may divide their subject matter areas differently from the way psychological abilities cohere (and multiple factor representations are possible, e.g., Maraun, 1996).

In addition to modeling abilities that make up personal intelligence, a second purpose of these studies is to learn how people's personal intelligence is distributed across the range of human ability. General intelligence is regarded as normally distributed, although there is some evidence that it is becoming negatively skewed (i.e., fewer people score below average) as nutrition improves worldwide and positive cognitive stimulation increases (e.g., Colom, Lluis-Font, & Andrés-Pueyo, 2005).

The distributions for people-focused intelligences appear far more pronounced in their negative skew, and we expect that to be the case here (cf., Maul, 2012). Although an individual with good reasoning ability can identify people's inner qualities, and predict people's consistent behavioral expressions, such predictions are likely to be limited by the sheer complexity of human behavior. At the lower-ability end of reasoning, however, a sizeable group of people may miss even the basics about themselves and others. This lower-scoring group may include some people who experience Autism-spectrum symptoms (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001).

We also examine the TOPI in relation to self-estimated personal intelligence, exploring whether ability and self-judged personal intelligence are distinct constructs. And, finally, we examine relationships between ability-based subscales of personal intelligence and earlier-studied criteria.

Studying a concept together with the way it is measured creates a virtuous cycle of better measurement and better theoretical understanding (Borsboom, Mellenbergh, & van Heerden, 2004; Hood, 2009). Test subscales are better justified when they are based on the actual structure of abilities measured by the test—as opposed to scales that are created conceptually by authors without regard to empirical independence among scales (Sinharay, Puhan, & Haberman, 2011). Moreover, understanding the abilities people use to solve problems in a domain can be helpful to improving their performance through education.

2. The TOPI series of tests

2.1. General overview of the TOPI tests

The *Test of Personal Intelligence (TOPI)* was developed to examine the viability of measuring a personal intelligence. As the scale underwent revisions from versions 1.0 to 1.2, evidence accrued for the construct (Mayer et al., 2012). The present version, 1.4, is a 93-item subset of the TOPI 1.2 (Mayer & Skimmyhorn, 2017). Most key characteristics of the TOPI have remained the same over forms. Across the test, all items are multiple choice, with four alternatives each. The test items fall within one of four areas of problem solving, reflecting the four theoretical areas from “Recognizing Information” to “Systematizing Plans.”

The four problem-solving areas of the TOPI further divide into 13 clusters of more specific test content and similar format. 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 (alternative “d”) must assess the given behaviors of being with people, going out with them, and consequent enjoyment, and extract from them the most likely motive; in this case, to socialize. The logic required for each task is different. For example, the Systematizing Plans area includes a cluster of items that ask whether test-takers can recognize goals that conflict (“problematic goals”); see Mayer et al. (2012) for more examples of clusters and their items.

The TOPI's theoretically-guided coverage ensures both that all its test questions address reasoning about personality—and that a broad range of relevant content areas are sampled, providing evidence for its adequately representative coverage, as recommended in the *Standards for Educational and Psychological Testing* (Joint_Committee, 2014).

2.2. Veridical scoring and item difficulty

Correct answers to TOPI items are scored 1 point; incorrect, 0 points. Correct answers were identified with reference to relevant, generally-agreed upon findings from published research in the field of personality psychology. Using only the most agreed-upon findings selects for items that may seem easy at times, but that are most apt to reflect valid knowledge about personality rather than agreement with one theorist or empirical article at the expense of another.

2.3. Self-judged personal intelligence

People also can be asked to estimate their own personal intelligence using items keyed to the theory, and we have created a scale that helps them do so called the *Self-Estimated Personal Intelligence* inventory (SEPI). Its items take the form, "I read people's intentions well" and "I don't know who I am" (reversed), and are answered from "strongly disagree" to "strongly agree" on a 5-point scale. The SEPI is unlikely to reflect actual personal intelligence because self-judgments are influenced by high (or low) self-esteem and are limited by 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). We also will test the correlation between the TOPI and SEPI.

3. Study 1. Factor analyses and item analyses of the TOPI 1.4

3.1. Purposes and hypotheses of the study

The key purpose of Study 1 was to explore the possible representations of distinct abilities that make up personal intelligence and—should they exist—to develop scales that measured them. We were guided by several hypotheses:

Hypothesis 1. *People's attention drops while taking the test*. Item analyses are better if people pay equal attention to all the items of the test. We therefore assessed drop-off in attention of test-takers by examining their endorsement rates of unlikely responses and their overuse of a single letter response (long-string responding) over the length of the test. We planned to screen out any participants who exhibited extreme declines from the first to second halves of the test or other indicators of inattention 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 $r = .55$ to .95, we expected that any subsidiary factors of personal intelligence would correlate in the upper portion of that range. Consistent with contemporary practice, we planned to retain any factor with <85% of its variance explained by overall personal intelligence (e.g., O'Connor Quinn, 2014). Estimates of such explained common variance (ECV) can be carried out by using ancillary bifactor models (Reise, 2012) alongside the simple-structure correlated factors that were the focus of our analyses.

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., exhibit configural invariance). We also expected that women would score somewhat higher than men on the test as they do on 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 intelligences such as personal and emotional intelligences (see Fiori et al., 2014; Mayer, Caruso, & Salovey, 2016). We expect nonetheless that a substantial number of people will exhibit poorer 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 with a measure of Self-Estimated Personal Intelligence (SEPI).

4. Method

4.1. Participants

Participants were 10,618 test-takers drawn from seven samples, mostly from the United States military. The overall 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 them can be found in the footnotes to Table 1.

A master data file was assembled from the seven samples, an ethnically diverse group who had been tested between fall 2012 and June 2015 by the United States *Office of Economic and Manpower Analysis* (OEMA), as well as in our laboratory. 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 combined sample of individuals tested while in in Officer Candidate School in June 2015. We chose June 2015 as our stopping point because by then our samples contained over 10,000 participants 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.

4.2. 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.

5. Results

5.1. Participant screening based on attention levels (Hypothesis 1).

All TOPI data provided by the OEMA had been prescreened to ensure 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 scale, counted the number of times a participant selected the *least endorsed* answer among the 10,618 respondents (the most obviously incorrect), for each of the 93 TOPI items. Infrequency scales for the first and second halves of the test correlated $r = .63$. The second, Letter-Repetition scale, 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 $r = .34$.

Table 1
Overview of the original and combined samples of study 1.

Sample number and description[a]	Sample sizes				Age	
	N	*Ns* after screening			Mean	Range
	Total	Total	Men	Women		
1. West Point Military Academy, Class of 2014[b]	1,114	1,106	905	201	20.98	20–30
2. West Point Military Academy, Class of 2015[b]	1,048	1,048	867	174	20	20–22
3. West Point Military Academy, Class of 2016[b]	972	968	807	161	20.52	20–22
4. West Point Military Academy, Class of 2018[b]	1,089	1,078	832	246	20.08	20–22
5. Reserve Officer Training Core (ROTC)[b]	5,614	5,390	4,208	1,176	20.01	20–30
6. Mturk sample of "half-time or more" employees, screened for inattentive responding (Lortie, 2015)[c]	474	459	210 [c]	247 [c]	34.91	19–70
7. United States Army Enlistees in Officer Candidate School; several smaller samples tested 2013–2015 combined[b]	307	276	220	56	25.80	22–90[e]
Overall total	10,618	10,318	8,049[d]	2,261[d]	21.09	19–90[e]

[a] Participants in all samples were required to answer every question on the TOPI.
[b] Members of samples 1–4, 5, and 7 were tested by the Office of Economic and Manpower Analysis (OEMA) of the United States Army. Examinees took the TOPI in large, proctored groups, 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 and that their scores would be used by the OEMA to help them choose the best branch of the U.S. Army to enter into once they had graduated (Mayer & Skimmyhorn, 2017). Military data were scored and examined under a data-sharing agreement between the OEMA and the [*Removed for anon. review*].
[c] Test-takers in the non-military sample, Sample 6, completed their tests online through MTurk; they each received small payments, contingent on their completion of all the questions. Sample 6 participants were permitted to specify a gender other than male or female; see Lortie (2015) for details.
[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."
[e] The two age responses of "90" in Sample 7 likely reflected inattentive or mischievous responses, or keystroke errors.

5.1.1. Overall responding and attention fall-off

The modal response rates of 0% for the infrequency scale and approximately 25% for Letter Repetition indicated that most participants responded meaningfully throughout the test; there was, however, a long, low-frequency tail of respondents who scored at much higher levels on each scale. A nonparametric one-sample Kolmogorov-Smirnov analysis of the mean difference in scales across the first- and second test halves found evidence for more infrequent responding and more letter repetitions during the second half of the test ($ps < .0001$).

5.1.2. 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 on the Infrequency scale if they selected the least frequently-endorsed alternatives ≥25% of the time (a random-like pattern), and a warning on the Repetition scale if they answered with a single letter >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. Given the evidence that some participants experienced a drop in attention, we also created parallel scales for the second half of the test alone, prorating the cut points for those versions. We then screened out respondents with one or more warnings or two or more cautions on any of the four scales; this excluded 300 (2.8%) of the 10,618 participants. Across the seven subsamples, exclusion rates generally ranged between .4% and 4%. The highest rate of removal (about 10%) was from the Officer Candidate School sample, perhaps owing to a lower commitment to the testing process among that group. Participants older than 25 years-of-age were excluded slightly more than those younger (.05% v. .02%), and men were excluded a bit more than women (.03% v. .02%).

5.2. 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 represent all samples equally and so the split could be readily re-created as needed.

5.3. Addressing the factor structure of personal intelligence (Hypothesis 2)

We next asked whether personal intelligence could 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.

5.3.1. Overview of the TOPI 1.4 factor analyses

We evaluated factor 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). TOPI items are dichotomously scored (i.e., correct or incorrect) and so categorical in form; we therefore used M*plus*'s (Version 7.2) weighted least squares, mean and variance adjusted estimation (WLSMV) as it is particularly appropriate for these data (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). We compared delta and theta parameterizations where they were both available (in confirmatory analyses) and they were essentially identical; we used the theta option so as to compare modification indices for item pairs.

5.3.2. 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 Table 2 labeled under the row "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": several items came from a single item cluster and shared wording, rather than representing a more meaningful construct.

Table 2
Factor models of the TOPI 1.4.

Model tested	Items deleted	Items & split	Variables/free parameters	Fit indices					$r_{factors}$
				Chi-2	df	RMSEA	CFI	TLI	
Initial exploratory factor analyses, odd sample (N = 5,144)									
1- to 3-factor solutions facparsim-rotated, oblique									
One factor model	0	na	93/93	9,813.11	4,185	.016	.907	.905	na
Two factor model	0	93: 43/50	93/185	7,298.51	4,093	.012	.947	.945	$r_{I,II} = .46$
Three factor model	0	93: 34/39/20	93	6,488.78	4,002	.011	.959	.956	$rs_{I\ to\ III} = .35$ to .52
Confirmatory two factor models, odd sample (N = 5,144)									
Removing cross-loading items (>.25 on both factors)									
Simple structure	13	80: 40/40	80/161	7,127.40	3,079	.016	.912	.910	$r_{I,II} = .79$
Removing low-loading items (<.25)									
Simple structure	21	72: 36/36	72/145	5,354.72	2,483	.015	.934	.932	$r_{I,II} = .80$
Final factor model, odd sample (N = 5,144)									
Removing items with poor pairwise fit (large modification indices)									
Simple structure	25	68: 34/34	68/137	4,074.63	2,209	.013	.952	.950	$r_{I,II} = .82$
Bifactor Model[a]	25	68: 68/34/34	68/204	3,408.78	2,142	.011	.967	.965	$rs_{I,II,ov} = .00$[c]
Configural (factor) invariance for two factors for women and men[b]									
Men, separately	25	68: 34/34	68/137	3,632.04	2,209	.013	.955	.953	$r_{I,II} = .82$
Women, separately	25	68: 34/34	68/137	2,454.66	2,209	.010	.948	.947	$r_{I,II} = .82$
Combined model	25	68: 34/34	69/208	5,722.63	4,484	.011	.962	.962	$r_{I,I,I,/m/f} = .82$ to .81
Confirmatory two factor models, even sample cross-check (N = 5,174)									
Simple structure	25	68: 34/34	68/137	4,226.71	2,209	.013	.947	.945	$r_{I,II} = .81$
Bifactor Model[a]	25	68: 68/34/34	68/204	3,380.06	2,142	.011	.967	.965	$rs_{I,II,ov} = .00$[c]
Study 2 cross validation sample (N = 8459)									
Simple structure	26	67: 33/34	67/135	5,682.48	2,143	.014	.957	.956	$r_{I,II} = .87$

[a] The bifactor model employs an overall factor on which all items load, and individual group factors (in this case, two). The overall and group factors are constrained to correlated zero with one another. Consequently, every item is a product of two independent sources of variance: one due to the overall factor (overall personal intelligence) and the second due to the group factors.
[b] This test of configural invariance indicated whether the number of factors fit for equally well for men and women. The model was first fit to groups of women and men separately, and then in a combined model including both groups.
[c] As noted in 'a', the correlation between the two specific factors, designated I and II, is constrained to zero in the bifactor model, as are their correlations with the overall factor (ov), which allows for estimating the variance accounted for by the overall factor relative to specific factors (see text).

5.3.3. Interpreting the factors

Both factors from the two-factor model appeared interpretable. We labeled Factor I "Consistency-Congruence" (CC) personal intelligence because its items asked about consistent patterns across traits. Twenty of its 33 items concerned understanding which socio-emotional traits go together (e.g., liveliness with talkativeness) and how mental states and desires reflect motivational patterns (e.g., talkativeness and liking people leads to socializing). Nine items reflected the ability to identify a trait's relations to behavioral outcomes (e.g., that careless people are apt to cause property damage), and how motivational memories might motivate specific behaviors. The four remaining items concerned identifying specific acts, behaviors, or other choices that might fulfill a need.

We labeled Factor II "Dynamic-Analytic" (DA) personal intelligence because it involved reasoning about personality dynamics and integrating information. Ten of its 34 items concerned recognizing problematic goals and goal conflicts (e.g., "to be able to please everyone"); another 10 concerned the ability to use personal memories to motivate oneself (e.g., "remembering a careless act that turned out badly so as to be more careful"), choosing actions that could improve oneself ("taking a course"), and understanding how behaving in a certain way could change one's self-perception ("acting smart"). Another eight items concerned making sense of the disagreements of observers about a person; a further five items concerned anticipating a person's future behavior (or reconstructing their past behavior) from their traits. (The final item involved trait understanding). Both factors drew items from all four problem-solving areas represented in the test.

5.3.4. First confirmatory test of the two-factor model

Given its interpretability, 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 earlier. 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) to check the scales' dimensionality. We calculated the amount of variance of each factor due to the common personal intelligence factor, termed its *explained common variance* (ECV, Reise, Moore, & Haviland, 2010). The overall personal intelligence factor accounted for 72% of the variance of the first factor and 79% of the second (76% overall). Anything <85% is regarded as supporting multidimensionality and therefore supported the use of two separate factors (O'Connor Quinn, 2014; Stout, 1990).

5.3.5. Configural invariance (Hypothesis 3)

We next 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.

5.3.6. Re-confirmation on the cross-check sample

As a further step, we examined the fit of the model in the cross-check sample ($N = 5174$). The fit of both the simple structure and bifactor models was almost identical to that found in the exploratory sample (Table 2).

5.4. Item analyses of the scales

We next conducted item analyses of the two scales individually in IRTPRO (Cai, Thissen, & du Toit, 2016). We employed a 2-parameter IRT model of each scale that estimated each item's difficulty level and power of discrimination, using 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 −2loglikelihoods across models, where $\chi^2 = -2ll_{3PL} - 2ll_{2PL}$. For the CC scale, these values were $\chi^2(34) = 141{,}095.85\text{–}141{,}066.26 = 29.59$, n.s., and for DA, $\chi^2(34) = 169{,}466.07\text{–}169{,}443.66 = 22.41$, 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).

5.4.1. 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-χ^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.

5.4.2. 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, for a total of 67 items.

5.4.3. Item difficulty and discrimination across ability levels

For the two scales, the item slopes (a parameters) ranged from 0.48 to 1.48, with a mean of .84 for CC, and from .39 to 1.70, with a mean of .87 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.

5.4.4. 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 (which varied across some 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.

5.4.5. 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 $M_{diff} = -.02$.

5.5. Characteristics of the final scales

5.5.1. Scaled scores

We created scaled scores using a T-scale (i.e., $M = 50$ and $S = 10$) for both the Consistency-Congruency and Dynamic-Analytic scales based on the cross-check sample.

Table 3
Means, reliabilities, and correlations for the TOPI 1.4R *Consistency-Congruence*, *Dynamic-Analytic* and *Composite* scales.

TOPI 14R scales	Descriptive statistics			Reliabilities		Correlations		
	Total mean (S)	Men[a] mean (S)	Women[a] mean (S)	Alpha[c]	Margin. (IRT)[e]	Consis.	Dyn.	Orig. TOPI 14
Study 1 cross-check (even-numbered) sample, $N = 5{,}174$								
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[b]	49.76 (7.38)	49.46 (7.45)	50.80 (8.49)	.85**[d]	Not est.[f]	.88**	.90**	.96**
Study 1 total sample, $N = 10{,}318$								
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[b]	49.88 (7.45)	49.53 (7.52)	51.13 (7.06)	.85**[d]	Not est.[f]	.89**	.90**	.97**
Study 2 total sample, $N = 8{,}459$								
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[b]	49.36 (7.99)	49.02 (8.10)	50.49 (7.48)	.88**[d]	Not est.[f]	.90**	.90**	.97**

*$p < .05$.; **$p < .01$.
[a] Men and women number 6,843 and 1,971 in the cross-check sample, 8,049 and 2,261 in the Study 1 total sample, and 6,539 and 1,920 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.
[b] Formed from the mean scaled scores of the Consistency-Congruency and Dynamic-Analytic factors (see text).
[c] Based on standardized items.
[d] The alpha reflects the reliability of the 67 items before scaling; the alpha for the scaled score cannot be calculated.
[e] 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.
[f] The marginal reliabilities of the simple summed score of two IRT-based scaled scores cannot be estimated.

5.5.2. Test and scale correlations

The top of Table 3 indicates an $r = 0.58$ correlation between the Consistency-Congruence and Dynamic-Analytic scales for the cross-check sample (the results for all participants are in the second portion of Table 3). A score that averages the two scaled scores (CC and DA), referred to as the TOPI 1.4R Composite, also is reported, along with the original TOPI 1.4 total (all 93 items).

5.5.3. 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: $SEM_{below} = 4.3$; $SEM_{above} = 5.9$).

5.5.4. Scale reliabilities (Hypothesis 5)

As implied by 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 ranged $\alpha = .74$ 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 shows that the same statistics for the full sample of Study 1 were similar to those of the cross-check sample.

5.6. Correlations between actual personal intelligence and self-estimated personal intelligence (Hypothesis 6)

We further had hypothesized that the *Test of Personal Intelligence* drew on mental abilities that were distinct from self-judgment. The master data file also included scores on a 16-item version of the Self-Estimated Personal Intelligence test (SEPI) for 8,866 cases. The SEPI-16 exhibited an alpha reliability of $\alpha = .87$ and correlated with the Consistency-Congruence and Dynamic-Analytic scales $r = .24$ and .24, and with the TOPI Total at $r = .26$, and the original 93 item TOPI 1.4, $r = .28$, $ps < .01$, levels slightly higher than our prediction. We will revisit the TOPI scales' validities in Study 3 and the General Discussion.

6. Study 2. Cross-validation in an independent sample

In Study 2 we reconfirmed the two-factor model of personal intelligence 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 (beginning several months after we had begun analyses of our master file). This second wave of data were scored for the OEMA but otherwise remained unexamined until we had completed the analyses for Study 1 in late September of 2016.

6.1. 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.

6.2. Measures, procedures, and screening

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

6.3. Results

6.3.1. Screening

In Study 2, the screening (unchanged from Study 1) resulted in the removal of 355 test-takers or 4%, a rate 1.2% higher than the 2.8% exclusion rate in Study1, leaving $N = 8,459$.

6.3.2. 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. The correlation between the two factors rose from $r = .82$ to .87 from Study 1 to 2, due in part to the greater range of test-taking ability in Study 2 (correcting the obtained correlation in Study 1 for range restriction increases the $r = .59$ halfway (to .614) to the obtained $r = .64$ of Study 2). The Explained Common Variances for the two factors (tested within a bifactor model) were 83% and 81%, which continued to argue for the presence of two factors. 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.

7. Study 3. Testing the TOPI 1.4R scales' correlations with criteria. A reanalysis of Mayer et al., 2012 (Study 3)

To understand more about the *Consistency-Congruence* and *Dynamic-Analytic* scales developed here (and their composite), we rescored the TOPI 1.2 used in Mayer et al. (2012, Study 3) for the subset of 67 items that now formed the Consistency-Congruency and Dynamic-Analytic scales of the TOPI 1.4R (all of which appeared in the original scale). In the earlier Study 3, we had correlated the TOPI with other psychological tests measuring intellectual ability, the Big Five, and several other psychological scales. In this reanalysis, we hypothesized that (a) the two factor-based scales would exhibit reliabilities and correlate with one another approximately as in Studies 1 and 2 and (b) that they would exhibit differential correlations with at least several criterion measures used in the original Study 3.

7.1. 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 additional measures described in the following "Criterion Correlations" section.

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

8. Results

8.1. 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.

8.2. Criterion correlations

Table 4 includes the correlations between the two subscales and criteria from the earlier study. Both scales and the total correlate significantly with 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 $r = .31$ to .64 range, $ps < .001$. They correlate with Agreeableness and Conscientiousness in the Big Five $r = .12$ to .20, with Psychological Mindedness $r = .29$ to .39, and negatively with symptoms of personality disorders $r = -.07$ to $-.22$. They show

Table 4
The TOPI 1.4R and criteria: reanalysis of data from Mayer et al., 2012 ($N = 384$).

TOPI and criterion measures	TOPI 14R factor scales[a]		Difference	Composite
	Consistency	Dynamic	(Absolute)[a,b]	Score[a]
TOPI scales and other mental abilities				
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 ability	.37**	.51**	.14**	.48**
Managing ability	.54**	.64**	.10**	.65**
Big Five				
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**
Psych minded-total	.29**	.39**	.10**	.38**
Discussing prob.	.25**	.33**	.08	.33**
Access feelings	.18**	.21**	.03	.22**
Understanding	.12	.15**	.03	.15**
Motivation	.24**	.29**	.05	.29**
Open change	.09	.15**	.06	.13
Symptomatology				
Maladaptive agreeableness	−.14**	−.15**	.01	−.16**
Narcissistic grandiosity	−.13	−.22**	.09	−.19**
Narcissistic personality (NPI)	−.07	−.13*	.08	−.11**
Self-described social skills				
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
Life space questions (Over the past…week did you:)				
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**

[a] Only correlations at $p < 0.01$ are indicated to control for Type I Error.
[b] Significance levels are a consequence of both difference and correlation level.
** $p < 0.01$.

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. Beyond that, there is just a hint (not quite $p < .01$) 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 in the General discussion.

9. General discussion

9.1. 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, 2017; Mayer et al., 2012). Facilitated by a now larger sample of test-takers, we here examined whether there existed subsidiary mental skills that made up personal intelligence, the distribution of the ability, and its relation to self-judged insight into personality.

We found evidence that people can be characterized as employing two closely-related abilities to solve problems in the area, Consistency-Congruence and Dynamic-Analytic reasoning, and that these were largely unrelated to how smart test-takers thought they were about people, as indicated by the *Self-Estimated Personal Intelligence (SEPI)* measure. 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 scales to the TOPI 1.4R to monitor attentive responding.

9.2. The two-factor model of the test—and a one-factor representation as an alternative

9.2.1. Consistency–Congruence and Dynamic-Analytic reasoning

The two-factor model employed here divides people's personal intelligence into Consistency-Congruence and Dynamic-Analytic personal intelligences. 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 different people can perceive one another differently, and how to set goals for the future.

9.2.2. A one factor alternative

Although the two-factor approach fits well, an alternative one-factor model also was viable, as supported both by the (slightly less-well fitting) one-factor model itself and by the bifactor model, which includes a common factor (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.

9.2.3. 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 correlate at estimated levels of between $r = .81$ and .87 in Studies 1 and 2 (assuming 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 within the range of similar estimates found for correlations among other broad intelligences. For example, 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 and Keith (1999) estimated correlations among broad abilities from $r = .61$ to .93.

9.3. 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: (a) identifying information, (b) modeling personality, (c) guiding choices and (d) systematizing goals. We regarded those problem-solving areas as likely distinct from the abilities people employ to think about solutions. Educators rightly distinguish between fiction and nonfiction when teaching literature, but their students use vocabulary skills and sentence comprehension to understand both. The parallel situation for personal intelligence is depicted in Fig. 1. There, the two reasoning areas, depicted in two horizontal arrow-like figures, cross-cut the four problem-solving areas (across the top). People may use Consistency-Congruence reasoning, for example, to identify trait-related clues in the Identifying Information area, or to select objectives consistent with one's traits in the Systematizing Goals area. Similarly, people may engage their Dynamic-Analytic reasoning to integrate contradictory information about a person's traits in the Identifying Information area, or to ensure that they choose mutually supportive goals where possible when Systematizing Plans.

9.4. 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 not assessed here. For example, Olson and Dweck (2009), have characterized some people as employing a *fixed* mindset in judging how people behave, emphasizing the stable nature of people's personalities, whereas other people favor a *growth*-oriented mindset that emphasizes the changing nature of personality. People who score higher on the Consistency-Congruence scale may favor the fixed mindset; those higher on the Dynamic-Analytic scale may be more sensitive to personal growth.

9.5. The fifth person in the room (and the distribution of personal intelligence)

More people score highly on the TOPI than low 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). 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 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 who fall 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 although they perform satisfactorily, we could fairly say they "don't seem to get" a good deal about personality. 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, although above chance, simply couldn't answer much correctly. Note that these lowest-scoring individuals were apparently still paying attention: They avoided the relatively improbable alternatives that are flagged by the Infrequency Index. Yet many in this group struggled, and despite their apparent efforts to do well, selected incorrect answers much of the time.

In our daily experiences with other people, in other words, among a group of any 10 people there will be (on average) two who excel at understanding others, and three more who generally "get" other people. Among the five remaining people are three who are rather slow to pick up on the nature of themselves and other people, and two who recognize very little, perhaps nothing, about the personalities of the people around them. Because every fifth person answers fewer than 65% of questions about personality correctly, we might refer to them as the "fifth person in the room." Their test performance suggests that they are frequently confused as to people's traits and dynamics, and may,

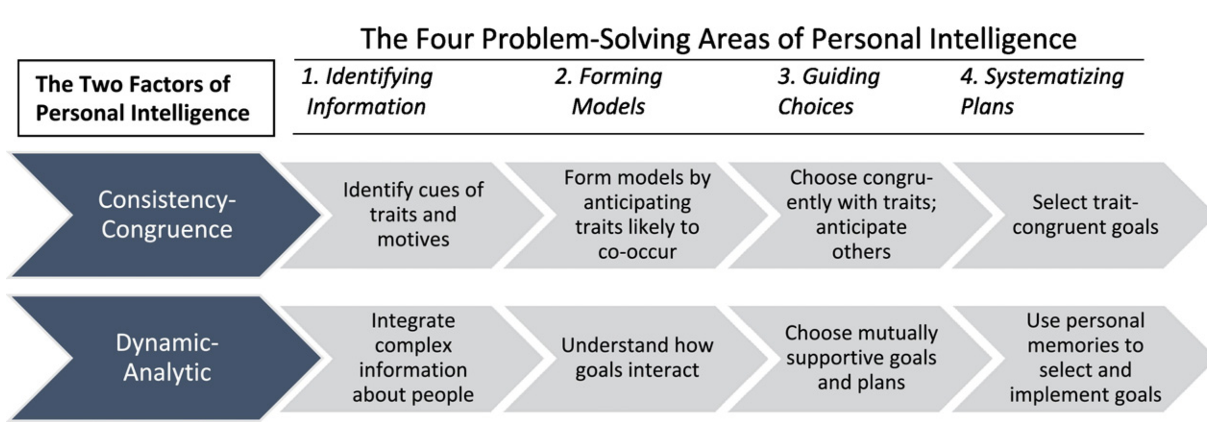

Fig. 1. Consistency and dynamic personal intelligences operate across the problem-solving areas of personal intelligence.

for example, mistake shyness for aloofness, attribute deceit to innocent forgetfulness, or pointlessly continue to criticize people who react to them with defensiveness. They also may be unaware of their misunderstandings: Self-estimated personal intelligence exhibited minimal relations with actual ability, 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" (Shakespeare, 1936/1972, 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 (Study 3).

9.6. 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 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 conflict. We believe that most people have sufficient ability in the area to benefit from curricular-based education, and to become better at such problem-solving should they so choose.

9.7. Future directions

There remains much to learn about personal intelligence, and the *Test of Personal Intelligence* can help. It has shortcomings to be sure, such as its relatively imprecise measure of high-ability test takers. There may also be other kinds of personally-intelligent problem-solving yet-to-be discovered. That said, 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. The roughly half of all test takers who appear to understand personality well must be very different from the below-average half 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 over time to affect a person's life course.

Acknowledgments

The authors often discussed statistical issues concerning this paper during its preparation. We are grateful that our conversations were enriched by Michelle Langer, Senior Statistician of the *American Institutes for Research,* who provided additional perspectives on the analyses. 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.

References

Appelbaum, S. A. (1973). Psychological-mindedness: Word, concept and essence. *The International Journal of Psychoanalysis*, *54*(1), 35–46.
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. http://dx.doi.org/10.1111/1469-7610.00715.
Boomsma, A., Hoyle, R. H., & Panter, A. T. (2012). In R. H. Hoyle, & R. H. Hoyle (Eds.), *The structural equation modeling research report* (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. http://dx.doi.org/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. http://dx.doi.org/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* (pp. 29–60) (3rd ed.). 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.
Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY US: Cambridge University Press.
Colom, R., Lluis-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. http://dx.doi.org/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. http://dx.doi.org/10.1002/job.1962.
DeYoung, C. G. (2015). Cybernetic big five theory. *Journal of Research in Personality*, *56*, 33–58. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1371/journal.pone.0098827.
Funder, D. C. (2001). Accuracy in personality judgment: Research and theory concerning an obvious question. In R. Hogan (Ed.), *Personality psychology in the workplace* (pp. 121–140). Washington, DC US: American Psychological Association. http://dx.doi.org/10.1037/10434-005.
Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York, NY US: Basic Books.
Hogan, R., & Foster, J. (2016). Rethinking personality. *International Journal of Personality Psychology*, *2*, 37–43.
Hood, S. B. (2009). Validity in psychological testing and scientific realism. *Theory & Psychology*, *19*(4), 451–473. http://dx.doi.org/10.1177/0959354309336320.
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. http://dx.doi.org/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) Durham, NH: University of New Hampshire.
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. *Emotion*, *14*(2), 358–374.
Maraun, M. D. (1996). The claims of factor analysis. *Multivariate Behavioral Research*, *31*(4), 673–689. http://dx.doi.org/10.1207/s15327906mbr3104_20.
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. http://dx.doi.org/10.1177/1754073912445811.
Mayer, J. D. (2008). Personal intelligence. *Imagination, Cognition and Personality*, *27*(3), 209–232. http://dx.doi.org/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. (2015). The personality systems framework: Current theory and development. *Journal of Research in Personality*. http://dx.doi.org/10.1016/j.jrp.2015.01.001.
Mayer, J. D., & Skimmyhorn, W. (2017). Personality attributes that predict performance of cadets at West Point. *Journal of Research in Personality*, *66*, 14–16.
Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer–Salovey–Caruso Emotional Intelligence Test (MSCEIT) user's manual.* Toronto, Ontario: Multi-Health Systems.
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. http://dx.doi.org/10.1080/00223891.2011.646108.
Mayer, J. D., Caruso, D. R., & Salovey, P. (2016). The ability model of emotional intelligence: Principles and updates. *Emotion Review*, *8*, 1–11.
McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10. http://dx.doi.org/10.1016/j.intell.2008.08.004.
Michell, J. (2012). Alfred Binet and the concept of heterogeneous orders. *Frontiers in Psychology*, 3. http://dx.doi.org/10.3389/fpsyg.2012.00261.
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.) Chapel Hill, NC: University of North Carolina.
Olson, K. R., & Dweck, C. S. (2009). Social cognitive development: A new look. *Child Development Perspectives*, *3*(1), 60–65. http://dx.doi.org/10.1111/j.1750-8606.2008.00078.x.
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. http://dx.doi.org/10.1111/1467-6494.00023.
Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. http://dx.doi.org/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. http://dx.doi.org/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. http://dx.doi.org/10.1080/00273170903504810.

Shakespeare, W. (1936/1972). *The complete works of William Shakespeare. Rockwell Kent illustrations.* Garden City, NJ: Garden City Books.

Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice*, *30*(3), 29–40. http://dx.doi.org/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. http://dx.doi.org/10.1007/BF02295289.

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. http://dx.doi.org/10.2307/1165261.