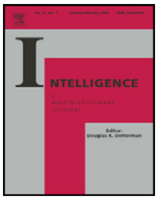
Intelligence 44 (2014) 93–102

Contents lists available at ScienceDirect

Intelligence



# Is health literacy an example of construct proliferation? A conceptual and empirical evaluation of its redundancy with general cognitive ability

Charlie L. Reeve *, Debra Basalik

*Health Psychology Program, University of North Carolina Charlotte, United States*

ARTICLE INFO

*Article history:*
Received 24 June 2013
Received in revised form 11 March 2014
Accepted 12 March 2014
Available online 16 April 2014

*Keywords:*
Health Literacy
Cognitive ability
Construct validity
Discriminant validity
Construct proliferation

ABSTRACT

During the last 20 years, health literacy has been promoted as an important determinant of individual and group differences in health outcomes. Despite a definition and pattern of associations with health outcomes highly similar to '*g*' (i.e., the general cognitive ability factor), health literacy has been conceptualized as a distinct construct. This study evaluates the conceptual and empirical distinctiveness of health literacy. A sample of 167 students from a southeastern urban university (117 females and 50 males) between the ages of 18 and 53 ($M = 21.31$, $SD = 5.61$) completed a cognitive ability battery, three health literacy tests, two knowledge tests, and a questionnaire assessing 12 health behaviors and health outcomes. Across 36 tests of criterion-related validity, cognitive ability had an effect in all 36 cases, where the health literacy tests only showed measureable incremental validity in 6 of 36 cases. Factor analysis revealed only three factors defined by the traditional ability tests with the health literacy measures loading on the ability factors as predicted by the content validity analysis. There was no evidence of a health literacy factor. The combined results from a comparative content analysis, an empirical factor analysis, and an incremental validity analysis cast doubt on the uniqueness of a health literacy construct. It is suggested that measures of health literacy are simply domain-specific contextualized measures of basic cognitive abilities. Implications for linking these literatures and increasing our understanding of the influence of intellectual factors on health are discussed.

## 1. Introduction

Health literacy has become a topic of significant interest among health and medical researchers during the past two decades, particularly in regard to its potential explanatory role in health disparities. For example, low health literacy has been shown to be associated with a variety of health outcomes including increased risk of chronic health problems and decreased utilization of health care services (Berkman et al., 2004). Findings such as these have compelled health researchers to consider the elucidatory role of cognitive factors with respect to individual and group differences in health outcomes. At the same time, researchers in the science of mental abilities have amassed evidence that individual differences in basic cognitive abilities, in particular *g* (the general mental ability construct), are associated with a variety of health behaviors and health outcomes. For example, it has been shown that measures of cognitive abilities predict health behaviors such as amount of physical activity, eating fruits and vegetables, taking vitamins, and smoking (e.g., Anstey, Low, Christensen, & Sachdev, 2009; Batty, Deary, Schoon, & Gale, 2007).

---

* Corresponding author at: Department of Psychology, University of North Carolina Charlotte, 9201 University City Boulevard, Charlotte, NC 28223-0001, United States.

*E-mail address:* clreeve@uncc.edu (C.L. Reeve).

http://dx.doi.org/10.1016/j.intell.2014.03.004
0160-2896/

Despite the apparent similarity between these streams of research, there has been little effort to understand how the health literacy construct fits within the broader nomological network of intelligence. In fact, it would appear that evidence demonstrating the influence of cognitive abilities on health outcomes is still relatively unknown health psychology, medicine and associated fields. To date, only a few studies have investigated how cognitive abilities and health literacy are associated (e.g., Mõttus et al., 2013; Murray, Johnson, Wolf, & Deary, 2011; Wolf et al., 2012). For example, Mottus et al recently showed that health literacy tests have somewhat limited incremental validity after accounting for cognitive ability and education. As such, it remains unclear to what extent health literacy measures are assessing construct variance that is distinct from basic cognitive abilities.

We believe this is a potentially important oversight as it creates the conditions for construct redundancy and construct proliferation. Construct proliferation and redundancy has been noted as a major problem in psychology (e.g., Le, Schmidt, Harter, & Lauver, 2010; Morrow, 1983; Schwab, 1980) and can be viewed as a major failure to adhere to the canon of parsimony in science. As these authors have noted, "a science that ignores the mandate for parsimony cannot advance its knowledge base and achieve cumulative knowledge." (Le et al., 2010, p. 112). We believe a clear understanding of the role of cognitive differences in health outcomes will emerge more quickly with an accurate understanding of the construct space being assessed by purported measures of health literacy. As such, our purpose is to evaluate the conceptual and empirical uniqueness of the health literacy construct compared to known cognitive abilities.

### *1.1. A conceptual analysis*

Within the larger process of construct conceptualization and validation, it is imperative to demonstrate the uniqueness of a construct by either its complete or partial independence from other comparable constructs (Cronbach & Meehl, 1955; Messick, 1989). The first step of that process requires the construct definition and nomological network of the target construct to be conceptually distinct from existing validated constructs. To avoid construct proliferation and redundancy, it is incumbent upon the newer proposed construct to distinguish itself from known constructs. In cases of where concepts compete for the same construct space, the parsimonious model is accepted as the superior model.

To begin, we can compare their definitions. Cognitive abilities (or "mental abilities") in general are defined as the sources of variance in performance on tasks requiring one to mentally process, comprehend, and manipulate information (Carroll, 1993; Reeve & Bonaccio, 2011a). While there are several specific abilities (e.g., quantitative reasoning; visual–spatial perception; cognitive speed), the general cognitive ability underlying all of these, or 'g' as it is known, has been defined as the factor reflecting individual differences in the ability to educe relations and correlations since it was first formally proposed by Jensen, 1998; Reeve & Bonaccio, 2011a; Spearman, 1904. As Gottfredson (2009) eloquently interpreted, this means "the generalized capacity to learn, reason and solve problems in essentially any context". Operationally, this means general ability would manifest behaviorally as the ability to obtain and understand information (i.e., to learn), process information (i.e., reason), and use information to make appropriate decisions (i.e., solve problems in context). Although the definition of health literacy used by researchers has evolved in the past fifteen years, the most commonly cited definition of health literacy is that adopted by the Institute of Medicine (Nielson-Bohlman, Panzer, & Kindig, 2004), the Department of Health and Human Services (Healthy People, 2010) and the World Health Organization (WHO) (1998) which states that health literacy is "the ability to obtain, process, and understand basic information and services need to make appropriate health decisions." Given the accepted definitions, it would appear that health literacy is essentially a restatement of the extant definition of 'g,' as it would be manifest in a specific context. That is, it is well established that general cognitive ability ('g') is indifferent to the indicator (Jensen, 1998; Spearman, 1927). This principle states that it is not the surface level features of items or situations that determine how well it measures 'g', but rather it is the underlying cognitive process. Any situation, regardless of context, that requires the eduction of relations and correlates and the application of knowledge to novel problem solving situations will measure 'g'. Thus, it appears that the primary conceptual difference is not in the cognitive processes per se, but in whether a model positing a general information processing capacity is more or less parsimonious than a componential model comprised of numerous domain-specific faculties.

In this respect, it has been predicted and confirmed that individual differences in health literacy scores are associated with health specific outcomes such as lower depression scores (Gazmararian, Baker, Parker, & Blazer, 2000), measures of morbidity, general health status, and use of health resources (DeWalt, Berkman, Sheridan, Lohr, & Pignone, 2004), mortality risk (Baker et al., 2007). However, in keeping with the definition *g*, Gottfredson (2004) posited that individual differences in *g* would manifest as the ability to comprehend, interpret and apply information in a health care context just as it has been shown to do so in educational, occupational, and social contexts (Kuncel, Hezlett, & Ones, 2004). This prediction has also been well supported. For example, evidence shows that individual differences in *g* are associated with lower depression scores, better general health, and significantly lower odds for stroke, congestive heart failure, chronic lung disease, heart problems, and hypertension later in life (Der, Batty, & Deary, 2009), reduced risk of mortality (Batty, Deary, & Gottfredson, 2008; Deary, Whalley, & Starr, 2003), reduced likelihood of smoking (Anstey et al., 2009; Reeve & Basalik, 2010), and higher diet quality and increased physical activity (Anstey et al., 2009; Batty et al., 2007) and other indicators of health (e.g., Mõttus et al., 2013). Thus again, this suggests potential for construct redundancy.

Thus, the conceptual question centers on the strength of the evidential basis for the most parsimonious theory. In this case, the empirical reality and the importance of the *g* construct have been well documented for more than 100 years. Going back to the early 1900s, a broad array of psychometric, biological, and behavioral genetic evidence (e.g., see Carroll, 1993; Jensen, 1998) has given rise to a broad consensus among experts that cognitive abilities are real psychological constructs (Gottfredson, 1997a; Reeve & Charles, 2008; Snyderman & Rothman, 1987), that they have a significant and meaningful influence on important real-world outcomes (e.g. Gottfredson, 1997b, 2004; Jensen, 1998; Kuncel et al., 2004; Schmidt & Hunter, 1998), that they have a basis in human biology and physiology (Deary, 2009; Gottfredson, 1997b [special issues of Intelligence]; Lubinski, 2004 [special section of the Journal of Personality and Social Psychology]) and is consistent with predictions from evolutionary psychology (Kanazawa, 2004, 2010). In contrast, the term "health literacy" did not appear until 1974 and a formal treatment of health literacy as a construct did not emerge until the late 1990s (Mancuso, 2009). Additionally, the proposal of a second domain-specific construct to replace a domain-independent construct is less parsimonious, and lacks a basis in the biology or physiology of the central nervous system. Moreover, it is not possible for a domain-specific adaptation of the brain to have evolved within the last century (arguably the time frame the domain in question has been in existence). Conceptually, it is difficult to see where or how health literacy might carve out a unique space within the existing network known as intelligence.

### 1.2. An empirical analysis

Recent research from IO psychology demonstrates that contextualized ability measures can yield incremental predictive validity above traditional ability measures despite denoting the same construct as the traditional measure (e.g., Klingner & Schuler, 2004). This suggests that contextualized ability measures may assess some unique variance in addition to the basic cognitive abilities that are relevant to context specific criteria. Thus, it remains possible that even if health literacy as a construct is not unique, the measures used do assess some unique cognitive skill that is distinct from general cognitive abilities. As such, we conducted an empirical examination of the convergent and discriminant validity of the three health literacy measures to better understand their empirical distinctiveness from basic abilities.

## 2. Methods

### 2.1. Participants

A total of 169 adults participated. Everyone enrolled in an introductory psychology course at a mid-size urban (largely commuter) university located in the southeastern U.S. were recruited through the university's undergraduate student research pool. Two participants did not complete the demographic background information. The remaining 167 participants (117 females and 50 male) were between the ages of 18 and 53 ($M = 21.31$, $SD = 5.61$), and mostly self-identified as white (62% white, 25.3% as African American and 5.4% as Hispanic). Most were single (91.9%), 4.7% were married, and 1 person was divorced. Similarly, most were childless (89.5%) and 8.1% had at least one child. About half were psychology or sociology majors (50.9%), 15.6% were business majors, 6.4% physical science majors, 6.4% nursing majors, the remainder spread among art, information technology, humanities and others. On a sing self-report item, 15.7% rated their family's financial status as "poor", 61.6% as "average" and 19.2% as "well-off." In terms of parental education levels, 45.2% indicated paternal education level as High School or less, 17.8% indicated an associates (or equivalent) degree, 22.3% indicated a bachelors degree, and 14.7% indicated a masters or terminal degree.

### 2.2. Measures

#### 2.2.1. Health literacy tests

The three most commonly used measures of health literacy are the *Rapid Estimate of Adult Literacy in Medicine* (REALM), the *Short Test of Functional Health Literacy in Adults* (S-TOFHLA), and the *Newest Vital Sign* (NVS). The REALM requires individuals to read out loud a list of sixty-six medical words arranged in order of increasing difficulty (Davis et al., 1993). The S-TOFHLA is a thirty-six item reading comprehension test and a seven item numerical ability test composed from materials encountered in a health care setting. The reading comprehension portion of the test measures an individual's ability to read and fill in missing words from a set of instructions that intended to prepare a patient for an upper gastrointestinal exam and a patient's rights and responsibilities portion of a Medicaid application form. The numerical portion of the test assesses an individual's ability to use numerical skills in the following of directions on a prescription bottle, interpretation of blood glucose levels, and honoring clinical appointments. The NVS presents the test-taker with an ice cream carton nutrition label and then asks six questions regarding the information presented in the label (Weiss et al., 2005).

#### 2.2.2. Cognitive ability battery

Six ability tests were selected for inclusion. Four scales from the Employee Aptitude Survey (EAS; Ruch, Stang, McKillip, & Dye, 2001) were used to assess *verbal reasoning*, *verbal comprehension*, *numerical ability* and *numerical reasoning*. The verbal reasoning test requires a person to combine separate pieces of information to form a conclusion. The verbal comprehension test requires a person to select the best synonym for a designated word from four alternatives presented. The numerical ability test measures the participant's ability to add, subtract, multiply and divide integers, decimals and fractions. The numerical reasoning test requires a person to analyze a series of numbers to determine the next number that would appear in the series. Alternate-forms reliability coefficients are reported to be .82, .85, .87 and .81 for the four scales, respectively (Ruch et al., 2001). Two-week interval test–retest estimates based on a similar sample reported by Reeve and Lam (2005) are .75, .74, .89, and .76, respectively. We also adapted the calendar task from the Kit of Factor Referenced Cognitive Tests (Ekstrom, French, Harman, & Derman, 1976). Participants are provided with a calendar, including dates and days of the week and asked a series of questions, such as "*What is the date of the second Monday of the month that begins on a Friday*?" This test matches the structure of the STOFHLA and NVS but independent of any health-related content. Finally, we included a reading fluency test specifically designed for this study to assess individual reading ability in a non-health related context, but using the same structure as the REALM.

Each word on the reading fluency test was matched to a word on the REALM in terms of number of letters, syllables, and frequency of occurrence in English language (as reported by WordCount.org).

### *2.3. Criterion measures (for the incremental validity assessment)*

#### *2.3.1. Health Knowledge Assessment (HKA)*

To assess general health knowledge we developed a multiple-choice test covering a wide range of health knowledge including aging, orthopedics and dermatology, children's health, chronic illnesses, common illnesses, mental health, nutrition and fitness, sexual and reproductive health, common medical terms and the human body. Full description of the HKA development is not reported her for space reasons but a report is available from the first author. The final version of the instrument included 60 items. Internal consistency for this sample was $\alpha = .74$.

#### *2.3.2. General information assessment*

An assessment of general knowledge was developed to assess the participant's general knowledge in contexts other than health. The purpose of including this as an outcome was to provide an additional test of discriminant validity for the health literacy tests (explained below). The 50 multiple-choice items assess a variety of domains including politics, geography, current events, history, literature, arts, science, grammar, sports and US trivia. Detailed information on its development is available from the first author.

#### *2.3.3. Body Mass Index (BMI)*

BMI was calculated from self-reported height and weight.

#### *2.3.4. Depressive symptoms*

We used the *Hopkins Symptom Checklist Depression Scale-13* (*HSCL-13*) which assesses the frequency of clinical symptoms of depression (Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 2007). Internal consistency for our sample was $\alpha = .83$.

#### *2.3.5. Self-rated overall health*

Participants rated their overall physical health on a 5-point Likert-type rating scale in response to the prompt, "*In general, compared to most people your age, how would you rate your overall physical health*?"

#### *2.3.6. Behavioral Risk Factor Surveillance System (BRFSS; CDC, 2010)*

We assessed several outcomes using items from the BRFSS. *Self-rated recent health* was measured with four items assessing the frequency during the last 30 days that the person felt sick, felt tired, felt extremely stressed out, or was unable to engage in usual activities due to poor mental or physical health. *Consumption of fresh food* assessed the frequency during the last 30 days that the person consumed at least one piece of fresh fruit or vegetable. *Unsafe dieting practices* asked how many days a month the person fasted for 24 h for the purpose of weight loss, took diet pills or other substance for weight loss, and took laxatives or vomited to lose weight. *Amount of exercise* asked how many days a month the person engaged in exercise for 30 min or more, and how many days they engaged in any physical activity. *Tobacco use* assessed how many days a month the person used tobacco products. *Alcohol use* was the sum of two items assessing how many days a month the person had (a) at least one alcoholic beverage and (b) five or more drinks in row. *Number of sexual partners* was assessed with a single self-report item.

### *2.4. Procedures*

Data were collected in groups of two or three. After informed consent, participants received a test booklet containing all instructions and measures. A trained research assistant proctored first proctored the administration of the ability tests and S-TOFHLA-Reading. After a short break, participants were directed to complete the remainder of the self-report measures at their own pace. During this time, a second research assistant accompanied each participant, one at a time, to another testing room to complete the oral and individual tests; namely, the S-TOFHLA-Numeracy, the calendar task, the NVS, the REALM and the reading fluency test.

## 3. Results

### *3.1. Descriptive analyses*

Descriptive statistics and uncorrected zero-order correlations among the ability and health literacy measures are shown in Table 1. Compared to the published norms for a general population of college students, the current sample scored, on average, at the 30th percentile on both the verbal comprehension and reasoning tests, and at the 15th percentile on both

**Table 1**
Means, standard deviations and zero-order correlations among cognitive ability and health literacy tests.

| | *M* | *SD* | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. Verbal comprehension | 16.64 | 4.37 | | | | | | | | |
| 2. Verbal reasoning | 14.31 | 4.88 | .29 | | | | | | | |
| 3. Reading fluency | 60.92 | 3.76 | .44 | .27 | | | | | | |
| 4. Numerical reasoning | 11.06 | 3.03 | .23 | .48 | .20 | | | | | |
| 5. Numerical ability | 30.63 | 11.21 | .23 | .41 | .31 | .60 | | | | |
| 6. Calendar task | 4.21 | 1.33 | .10 | .17 | .28 | .19 | .27 | | | |
| 7. REALM | 63.13 | 2.33 | .39 | .29 | .76 | .15 | .28 | .25 | | |
| 8. STOFHLA | 95.47 | 5.44 | .11 | .35 | .26 | .21 | .25 | .17 | .32 | |
| 9. NVS | 4.92 | 1.19 | .21 | .31 | .25 | .31 | .39 | .27 | .24 | .21 |

*Note*. $N = 168$. Correlations larger than |.16| are significant at $p < .05$; |.20| at $p < .01$.

numerical tests. Though low, these scores are consistent with the SAT scores of the local population of psychology majors typically attending the institution at which data were collected, and are consistent with performance by other samples drawn from this population of students on other ability tests (e.g., Bonaccio, Reeve, & Winford, 2012). Although the sample is somewhat homogeneous with respect to some cultural and background variables and has lower than average ability, the observed standard deviations indicate a normal amount of variance. All four tests had observed SDs in this sample similar to published norms (Ruch et al., 2001). This suggests that there may not be a significant concern about restriction in range in ability.

Because health literacy tests are typically designed for use in populations thought to be low in basic literacy skills, these tests tend to be comprised of items with low item difficulty parameters. As such, ceiling effects could be of concern in our sample of a general population. And in fact, the means on all four literacy tests are high. The averages are all in the "adequate" classification range for each scale. Examination of the proportion of the sample scoring the maximum score reveals that only 5.8% of the sample scored the maximum on the REALM, 28% on the STOFHLA and 40% on the NVS. At the same time, there was sufficient variability. For example, 13.7% of our sample scored below the cut-off for "adequate literacy" on both the REALM and NVS. Nonetheless, test for non-normality show all three of these distributions are negatively skewed. The potential effect that these ceiling effects or skewness may have on the analyses reported below should be kept in mind.

### 3.2. Incremental validity analysis

To assess the degree to which the health literacy measures capture variance uniquely associated with health outcomes independent of basic cognitive abilities, we first conducted a series of hierarchical multiple regression analyses. For each outcome, we first regressed the outcome onto the controls of gender (coded as female = 0 and male = 1) and race (two dummy variables dichotomously coded for white vs. not and black vs. not). The variance accounted for by the controls is shown in the first column in Table 2. In the second step, scores from the cognitive ability battery were added to the model. The increment in $R^2$ due to addition of the ability tests is reported in the second column Table 2. [Note, the reason we entered the set of ability tests and not just an extracted *g*-factor is because our question concerns the conceptual and empirical distinctiveness of health literacy tests from basic cognitive abilities. Entering only g in step two would leave variance due to the narrow abilities to be claimed by the HL tests.] In the third step, one of the health literacy tests was entered to assess its incremental validity beyond cognitive ability. The third step was done three times, using each of the three health literacy tests. The increment in $R^2$ and the standardized beta weight for each health literacy test are reported in the remaining columns in Table 2.

The set of cognitive ability tests account for additional variance in all outcomes beyond the controls. In fact, given that an $R^2$ of .05 is equivalent to a medium effect size (Rosenthal & Rosnow, 1991), cognitive ability has at least a medium effect (or larger) on seven of the 12 outcomes and the equivalent of a small effect for the other 5 outcomes. In all cases, cognitive ability is positively associated with better outcomes. This pattern of criterion-related validity coefficients is consistent with the bulk of the literature in cognitive epidemiology showing positive associations between cognitive ability and positive health outcomes and inverse associations with negative outcomes.

In contrast, out of the 12 outcomes tested, the REALM only accounted for any measurable amount of incremental variance in four cases, the largest of which would be classified as a small effect (Unsafe dieting behaviors). However, in this case higher scores on the REALM were paradoxically associated with increased rates of unsafe dieting behaviors. The S-TOFHLA only explained a statistically or practically significant (i.e., at least a small effect size) increment in variance in 2 out of 12 outcome variables (Unsafe dieting behaviors and Alcohol usage). In this case, S-TOFHLA scores were associated with the outcomes in the direction expected (i.e., higher health literacy associated with fewer unhealthy behaviors). The NVS

**Table 2**
Analysis of incremental validity of health literacy tests over cognitive abilities.

| | Controls | Cog. ability tests | REALM | | STOFHLA | | NVS | |
|---|---|---|---|---|---|---|---|---|
| Outcomes | $R^2$ | $\Delta R^2$ | $\beta$ | $\Delta R^2$ | $\beta$ | $\Delta R^2$ | $\beta$ | $\Delta R^2$ |
| General information test | .15[a] | .28[b] | .04 | .00 | .03 | .00 | .11 | .01 |
| Health knowledge test | .02 | .15[b] | .08 | .00 | .00 | .00 | −.04 | .00 |
| Self-rated overall health | .03 | .03 | .08 | .00 | −.04 | .00 | −.17 | .02 |
| Self-rated recent health | .08[a] | .04 | .06 | .00 | .03 | .00 | −.04 | .00 |
| Fresh fruit/veggies | .05 | .09 | −.05 | .00 | .15 | .02 | .08 | .00 |
| Unsafe dieting practices | .02 | .03 | .28[a] | .04[a] | −.28[b] | .06[b] | .07 | .00 |
| Amount of exercise | .04 | .05 | −.04 | .00 | .04 | .00 | −.05 | .00 |
| BMI | .10[b] | .05 | −.12 | .01 | −.11 | .01 | .20 | .03 |
| Depressive symptoms | .05 | .02 | .10 | .00 | .09 | .01 | .01 | .01 |
| Number of sexual partners | .02 | .05 | .07 | .00 | −.10 | .01 | .20 | .03 |
| Tobacco usage | .04 | .07 | .12 | .01 | .04 | .00 | .12 | .01 |
| Alcohol usage | .01 | .03 | .19 | .02 | −.25[a] | .05[a] | .08 | .01 |

*Note*. Listwise deletion results in $N = 150$. $\beta$ = standardized regression beta weight. BMI = Body Mass Index. Controls entered at first step include gender (female = 0; male = 1) and race (using two dummy codes where white = 1 & non-white = 0, and black = 1 & non-black = 0). The set of ability tests were entered in second step. Only the health literacy test indicated was entered at step three.

[a] Indicates $p < .05$.

[b] Indicates $p < .01$.

did not obtain any statistically significant increments in variance across the 12 outcomes although it did add 3% of the explained variance in two cases (BMI and number of sexual partners). However, once again in both cases it was associated with the outcome in the wrong direction (i.e., higher health literacy was associated with higher BMI and high number of sexual partners).

An argument could be made that by including each HL test separately, we are not taking advantage of their aggregate test-specific variance. Thus, we repeated the incremental validity tests in two ways. First, we created a common Health Literacy factor by extracting the first principal component of the four HL tests. Then we repeated the regression analyses entering this factor score rather than a single HL test. In no case did the use of the factor score increment the variance explained more than did any one of the tests alone. Second, we entered all four HL tests as a battery in the final step (rather than computing a combined score) to allow for the maximization of test-specific variance. As expected, in no case did this result in an increment greater than that seen by simple addition of the increment due to the individual tests.

As an additional test of convergent and discriminant validity we evaluated the pattern of correlations with the health knowledge test and general information test. Conceptually, health literacy tests they should better predict scores on the health knowledge test than a general knowledge test, and should do so better than cognitive abilities. As shown in Table 3, this does not happen. All but one of the ability tests are significantly correlated with the health knowledge test whereas only the REALM correlates with it, but no more so than reading fluency or verbal comprehension. Second, all of the health literacy tests mimic the pattern shown by all of the ability tests in that they all correlate more strongly with the general information test than the health knowledge test. This inability to display a discriminant pattern of correlations further disconfirms the construct validity of health literacy.

### 3.3. Factor analysis

The battery of six ability tests and the four-health literacy measures (the reading and numeracy portion of the STOFHLA were scored separately) were submitted to an exploratory principal axis factor analysis with promax rotation (this was repeated with an orthogonal rotation, but it did not appreciably alter the lambda matrix). The KMO value was .76. Three factors with eigenvalues greater than 1.0 emerged (3.44, 1.34, and 1.11, respectively), accounting for 59.12% of the variance. The rotated factor-loading pattern is shown in Table 4. Again, the primary hypothesis being tested here is that the health literacy measures will denote a common factor distinct from those denoted by the ability tests. The lambda matrix shown in Table 4 does not support this hypothesis.

The first factor is dominated by the logical reasoning tests, but has minimal loadings from the STOFHLA Reading test and the NVS tests as well. Based on comparisons to well-validated models of cognitive abilities (McGrew, 1997, 2009; Reeve & Bonaccio, 2011a), this pattern of loadings strongly indicates the first factor reflects general reasoning ability (i.e., *g* or *Gf*). The second factor appears to be a basic verbal ability factor (e.g., *Grw*), possibly even reflecting crystallized vocabulary (e.g., *Gc*) given the dominance of the reading fluency and REALM tests. The third factor comprises STOFHLA-Numeracy, the calendar task and the NVS. At first glance it might appear to reflect a numeracy factor as the three tests with primary loadings do have numerical aspects. However, the fact that the two primary indicators of numerical ability loaded on a separate factor indicate this third factor reflects common aspects of these three tests other than the reasoning requirement. The only common component to all three is the need for visual scanning and search for information (e.g., *Gvs*). Thus, the set of tests appear to reflect a fluid reasoning factor, a lexical knowledge factor, and a visual-scanning factor.

Although there is little evidence from the factor analysis supporting the hypothesis that health literacy tests assess a unique source of variance, the communality estimates (shown in Table 3) show that the health literacy tests have a significant amount of test specific variance. For example, 75% of the variance in the NVS and 92% of the variance in the S-TOFHLA-Reading remain unaccounted for by this factor analysis. Although the fact the set of health literacy tests do not share a common source of variance among themselves beyond that attributable to cognitive abilities rules out the argument of a unique construct, it remains possible that

**Table 3**
Zero-order correlations of health literacy and cognitive ability tests with knowledge tests.

| | Health knowledge | General info test |
|---|---|---|
| *Health literacy tests* | | |
| REALM | .26[b] | .44[b] |
| STOFHLA | .07 | .14 |
| NVS | .10 | .30[b] |
| *Cognitive ability tests* | | |
| Reading fluency | .27[b] | .42[b] |
| Verbal comprehension | .33[b] | .54[b] |
| Verbal reasoning | .16[a] | .32[b] |
| Numerical reasoning | .17[a] | .28[b] |
| Numerical ability | .17[a] | .34[b] |
| Calendar task | .06 | .09 |

*Note*. $N = 168$.
[a] Indicates $p < .05$.
[b] Indicates $p < .01$.

**Table 4**
Results of exploratory factor analysis (principal axis factors).

| | Factor | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | Communalities |
| Numerical reasoning | **.88** | −.07 | −.09 | .67 |
| Numerical ability | **.69** | .04 | .06 | .54 |
| Verbal reasoning | **.51** | .06 | .15 | .40 |
| NVS | **.31** | .04 | **.30** | .25 |
| STOFHLA-Reading | .25 | .14 | −.10 | .08 |
| STOFHLA-Numeracy | −.08 | −.06 | **.68** | .38 |
| Calendar task | .07 | .04 | **.41** | .22 |
| REALM | −.07 | **.79** | .15 | .72 |
| Reading fluency | .02 | **.96** | −.07 | .87 |
| Verbal comprehension | .22 | **.45** | −.09 | .27 |

*Note*. $N = 168$. Promax rotation with Kaiser normalization. Salient loadings (>.30) shown in bold.

some of the remaining test specific variance in the health literacy tests may be uniquely associated with health outcomes. However, the prior incremental validity analysis confirmed that was not achieved.

## 4. Discussion

The purpose of this study was to evaluate the uniqueness of the health literacy construct and the validity of health literacy measures relative to cognitive abilities. Based on the totality of the evidence presented here, it is difficult to conclude that “health literacy” reflects a unique construct or that so-called health-literacy tests are psychometrically adequate measures. This conclusion is based on several lines of evidence. First, our conceptual analysis of the definition of health literacy found it to be largely synonymous with the established definition of general cognitive ability, and conceptually less parsimonious. Recent attempts to better define health literacy do not appear to have solved the problem of conceptual distinctiveness. Baker (2006) has recently promoted a framework for understanding health literacy that states it is comprised of an individual's reading fluency, vocabulary, arithmetic ability, and prior knowledge. This formulation of health literacy seems to posit health literacy as a manifestation of a set of cognitive abilities in the context of health related information. Even though this revised definition includes health knowledge as part the construct (rather than knowledge being an outcome of the construct as has been the case in the past), it does not set it apart from existing constructs within the intelligence domain. Baker's revised description of health literacy appears to align with the concept of a trait constellation: that is, it is cognitive capacity borne out in the application of reasoning abilities and acquired domain-specific knowledge to task specific demands. While this does relieve some of the redundancy with abilities alone, it does not appear to set it apart from domain-specific constructs within the existing framework of intelligence (see Reeve & Bonaccio, 2011a, for a detailed explication of the structure of intelligence).

Second, we empirically evaluated the incremental validity of health literacy measures net the influence of cognitive abilities. If health literacy measures assess a unique ability, they should be able to predict health outcomes, and in particular health knowledge scores, better than the traditional ability tests. Alternatively, it is possible that the real value of health literacy measures is that they are more effective, context-sensitive measures of cognitive abilities. Such contextualized measurement has been of value in applied areas such as employee selection (e.g., Klingner & Schuler, 2004). In either case, both predict health literacy measures to show incremental validity net traditional ability tests. However, consistent with the findings of Mõttus et al. (2013), our results disconfirm the suggestion that health literacy measures provide meaningful amounts of incremental validity relative to that provided by traditional (and more well-validated) measures of ability.

Third, results of a factor analysis showed that the health literacy measures loaded onto factors dominated by well-defined ability tests and did not form a unique factor. In fact, examination of the content the three most commonly used measures of health literacy show them to have content highly similar, and in some cases identical, to existing well-validated cognitive ability tests. For example, the REALM uses the exact same format as two scales on the Kaufman Test of Educational Achievement Second Edition (KTEA-II) and two subtests of the Peabody Individual Achievement Test — Revised (PIAT-R). All of these including the REALM require examinees to pronounce words of increasing difficulty that are presented without context. Likewise, the reading section of S-TOFHLA consists of two prose passages presented in a modified cloze procedure. The cloze method is a popular method of assessing reading comprehension (Carroll, 1993); a method used in the WJ-R (Neddenriep, Hale, Skinner, Hawkins, & Winn, 2007) and the Woodcock–McGrew–Werder Mini-Battery of Achievement (MBA; Flanagan et al., 1997). These results are consistent with others who have reported strong correlations between health literacy tests and traditional measures of fluid and crystalized abilities (e.g., Mõttus et al., 2013; Wolf et al., 2012).

Applying Carnap's (1950) total evidence rule, it is difficult to find support for the hypothesis that health literacy is a unique construct, and/or that health literacy tests measure something unique relative to traditional cognitive abilities. The combination of the evidence from across the conceptual and empirical analyses demonstrated that (a) the association between existing HL measures and health outcomes is likely a function of the degree to which the HL measures assess basic abilities, and (b) whatever causes the unique variance that is captured by HL measures, it is mostly unrelated to the health outcomes assessed. Similar to the conclusion reached by Mõttus et al. (2013), health literacy scores appears to be associated with health outcomes largely because they denote basic cognitive abilities.

It is important to note several limitations and concerns, however, before discussing potential implications of such conclusions. First, as noted previously, there is concern about the skewed nature of the distribution of scores on the health literacy tests. Because they were designed to detect illiteracy, they do often show ceiling effects when used in general populations. Ceiling effects can skew distributions significantly, and lead to concerns about attenuated correlations. This is a distinct concern for our data on the NVS and to some degree on the STOFHLA (but was not a concern on the REALM). This could distort both the incremental validity analysis and the factor analysis. In addition, we had access to a relatively limited sample size. This creates some statistical concern about somewhat larger sampling error in our coefficients. Also, many of our outcome measures were self-report. This leads to some concerns about the validity and reliability of the criteria used in the incremental validity analyses. While this might attenuate the magnitude of validity coefficients, it would not differentially affect ability tests and health literacy tests.

A broad look at our results may also lead to another conclusion that ironically poses a limitation to our data. Namely, our results suggest the current health literacy measures may not be psychometrically adequate for research purposes. This is a chilling implication given the apparent widespread use of these measures. From a methodological perspective, it suggests that proponents of a unique health-literacy construct have not yet developed a measure that can adequately assess the purported construct, which would disallow any formal study of the construct. From an applied perspective, it suggests the information being collected by practitioners may not be as valuable as once thought.

Despite these important limitations, the evidence from this study suggests that health literacy tests appear to be a case of “a sheep in fox's clothing”. We believe this has several important implications. First, rather than discrediting any prior evidence, we believe this general finding provides a great opportunity for both health researchers and psychologists. Evidence such as this allows researchers to merge these two bodies of evidence. This gives medical researchers access to a century's worth of accumulated data in the intelligence literature. Likewise, it allows those working in cognitive epidemiology to access the health literacy literature and add this evidence to its database. In short, our findings, combined with Mõttus et al. (2013), suggest health literacy researchers can reap the benefits of the intelligence literature with its extensive history, well-validated models and frameworks, and advanced quantitative and measurement methods. In turn, differential psychologists can reap benefits from the immediate infusion of data from health-specific contexts and populations that may not have otherwise been accessible. The amalgamation of these two literatures will likely yield an increased understanding of the role of intellectual factors that give rise to individual and group differences in health behaviors and health outcomes.

Second, the finding that health literacy is a reification of basic abilities has important implications for grounding expectations regarding attempts and interventions to increase “health literacy,” something that is a stated goal of several health organizations. It has been well documented that *g* is quite stable from adolescence through middle to late adulthood (see Reeve & Bonaccio, 2011b) and that programs designed to increase cognitive abilities, such as the Milwaukee Project or Head Start, do not have a significant lasting effect on ability test scores (e.g., Gilhousen, Allen, Lasater, Farrell, & Reynolds, 1990; Jensen, 1989; Locurto, 1991; Page, 1986; Scarr, 1993; Spitz, 1986). This implies that simple or short-term interventions to teach the abilities that are actually measured by “health literacy tests” may be unsuccessful.

On the other hand, there are numerous examples of programs that have successfully increased relevant domain-specific knowledge and procedural skills (e.g., Sternberg, Totff, & Grigorenko, 2009; Williams et al., 2002) even when these specific skills are highly correlated with ‘*g*’. Similarly, interventions targeted to increase health-related knowledge and teach specific behaviors have in fact demonstrated success. For example, among diabetic patients with low health literacy, utilization of both pictograms and teach back methodology significantly improved both their adherence to medication and dietary recommendations (Reza, Hassan, Hayedeh, Ramin, & Elham, 2012). Further, among patients with coronary artery disease, a clinical education intervention positively impacted the health behaviors and knowledge of both high and low literacy patients (Eckman et al., 2012). Thus, targeted interventions focused on increasing relevant knowledge and behavioral modification are more likely to improve the quality of life of those dealing with specific health related concerns. Critically, our findings show that existing health literacy tests do not measure criterion-relevant domain-specific knowledge or skills. To the extent they are associated with outcomes, it appears to be due to the fact they are short tests of basic abilities. Thus, we encourage health professionals to use valid assessments of criterion-relevant knowledge and skills, especially when the interest is interventions aimed towards the “patient”.

These findings also imply that health care professionals and providers should take seriously the cognitive demands of specific healthcare treatments or regimens. Our results again demonstrate that basic cognitive abilities are the underlying explanation for the association between “Health Literacy” tests and health outcomes. Because abilities are relatively resistant to change, our results imply that health care professionals should seek ways to provide important healthcare information that does not make large cognitive demands. Providing instructions in clear, easy to read and comprehend language would likely increase medication and treatment adherence, reduce return visits to hospitals, and increase treatment effectiveness. Similarly, providing easy to understanding health information would increase the odds that individuals would utilize such information in their day-to-day decisions that influence health (e.g., eating behaviors, exercise decisions, etc.). Interventions such as this are likely to be, relatively speaking, an inexpensive way to make immediate impact.

## 5. Conclusion

The current topic provides a critical demonstration of the need for all areas of social science to better heed the principles of measurement. As we noted earlier, posting new constructs similar to existing ones (the “old wine in new wineskins” phenomenon) can be viewed as a major failure to adhere to the canon of parsimony in science. If the mandate of parsimony were ignored, it would seem unlikely that we will achieve the broad scientific goal of uncovering the fundamental causes that underlie the apparent complexity of observed phenomenon (Toulmin, 1961; Le et al., 2010). As Le et al. (2010) stated, if we are to expect progress in research, “it is critical that we resolve questions of construct redundancy and construct proliferation and move in the direction of the ideal of scientific parsimony.” (p. 124).

## References

Anstey, K. J., Low, L. -F., Christensen, H., & Sachdev, P. (2009). Level of cognitive performance as a correlate and predictor of health behaviors that protect against cognitive decline in late life: The path through life study. *Intelligence*, *37*, 600–606.

Baker, D. W. (2006). The meaning and the measure of Health Literacy. *Journal of General Internal Medicine*, *21*, 878–883.

Baker, D. W., Wolf, M. S., Feinglass, J., Thompson, J. A., Gazmararian, J. A., & Huang, J. (2007). Health literacy and mortality among elderly persons. *Archives of Internal Medicine*, *167*, 1503.

Batty, G., Deary, I., & Gottfredson, L. (2008). Premorbid (early life) IQ and later mortality risk: Systematic review. *Annals of Epidemiology*, *17*, 278–288.

Batty, G. D., Deary, I. J., Schoon, I., & Gale, C. R. (2007). Mental ability across childhood in relation to risk factors for premature mortality in adult life: the 1970 British cohort study. *Journal of Epidemiology and Community Health*, *61*, 997–1003.

Berkman, N. D., DeWalt, D. A., Pignone, M. P., Sherida, S. L., Lohr, K., Lux, L., et al. (2004). *Literacy and health outcomes. Evidence report/Technology assessment No. 87 (RTI International–University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016). AHRQ Publication No. 04-E007-2.* Rockville, MD Agency for Healthcare Research and Quality.

Bonaccio, S., Reeve, C. L., & Winford, E. (2012). Text anxiety on cognitive ability tests can result in differential predictive validity of academic performance. *Personality and Individual Differences*, *52*, 497–502.

Carnap (1950). *Logical foundations of probability.* Chicago, US University of Chicago Press.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York Cambridge University Press.

Centers for Disease Control and Prevention (2010). *Behavioral Risk Factor Surveillance System Survey Questionnaire.* Atlanta, Georgia U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302.

Davis, T., Long, S., Jackson, R., Mayeaux, E., George, R., Murphy, P., et al. (1993). Rapid estimate of adult literacy in medicine: A shortened screening instrument. *Family Medicine*, *25*, 391–395.

Deary, I. (Ed.). (2009). *Intelligence, health and death: The emerging field of cognitive epidemiology. Intelligence*, *37(6)*, (Special Issue).

Deary, I. J., Whalley, L. J., & Starr, J. M. (2003). IQ at age 11 and longevity: Results from a follow up of the Scottish Mental Survey 1932. In C. Finch, J. -M. Robine, & Y. Christen (Eds.), *Brain and longevity: Perspectives in longevity* (pp. 153–164). Berlin Springer.

Der, G., Batty, G., & Deary, I. J. (2009). The association between IQ in adolescence and a range of health outcomes at 40 in the 1979 US National Longitudinal Study of Youth. *Intelligence*, *37*, 573–580.

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (2007). The Hopkins Symptom Checklist (HSCL): A self-report symptom inventory. *Behavioral Science*, *19*(1), 1–15.

DeWalt, D. A., Berkman, N. D., Sheridan, S., Lohr, K. N., & Pignone, M. P. (2004). Literacy and health outcomes. *Journal of General Internal Medicine*, *19*, 1228–1239.

Eckman, M., Wise, R., Leonard, A., Dixon, E., Burrows, C., Khan, F., et al. (2012). Impact of health literacy on outcomes and effectiveness of an educational intervention in patients with chronic diseases. *Patient Education and Counseling*, *87*, 143–151.

Ekstrom, R. B., French, J. W., Harman, H. H., & Derman, D. (1976). *Manual for kit of factor-referenced cognitive tests.* Princeton, NJ Educational Testing Service.

Flanagan, D. P., McGrew, K. S., Abramowitz, E., Lehner, L., Untiedt, S., Berger, D., et al. (1997). Improvement in academic screening instruments? A concurrent validity investigation of the K-FAST, MBA, and WRAT-3. *Journal of Psychoeducational Assessment*, *15*(2), 99–112.

Gazmararian, J., Baker, D., Parker, R., & Blazer, D. G. (2000). A multivariate analysis of factors associated with depression: Evaluating the role of health literacy as a potential contributor. *Archives of Internal Medicine*, *160*, 3307.

Gilhousen, M. R., Allen, L., Lasater, L. M., Farrell, D. M., & Reynolds, C. R. (1990). Veracity and vicissitude: A critical look at the Milwaukee Project. *Journal of School Psychology*, *28*, 285–299.

Gottfredson, L. (1997a). Editorial: Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*, 13–23.

Gottfredson, L. (1997b). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.

Gottfredson, L. (2004). Intelligence: Is it the epidemiologists' elusive fundamental cause of social class inequalities in health? *Journal of Personality and Social Psychology*, *86*, 174–199.

Gottfredson, L. (2009). Logical fallacies used to dismiss the evidence on intelligence testing. In R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing* (pp. 11–65).

Healthy People (2010). *U.S. Department of Health and Human Services, Office of Disease Prevention and Health Promotion.*

Jensen, A. R. (1989). Raising IQ without increasing g? A review of The Milwaukee Project: Preventing mental retardation in children at risk by H.L. Garber. *Developmental Review*, *9*, 234–258.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT Praeger.

Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. *Psychological Review*, *111*(2), 512–523.

Kanazawa, S. (2010). Evolutionary psychology and intelligence research. *American Psychologist*, *65*(4), 279–289.

Klingner, Y., & Schuler, H. (2004). Improving participants' evaluations while maintaining validity by a work sample — Intelligence test hybrid. *International Journal of Selection and Assessment*, *12*, 120–134.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*, 148–161.

Le, H., Schmidt, F. L., Harter, J. K., & Lauver, K. J. (2010). The problem of empirical redundancy of constructs in organizational research: An empirical investigation. *Organizational Behavior and Human Decision Processes*, *112*, 112–125.

Locurto, C. (1991). Beyond IQ in preschool programs? *Intelligence*, *15*, 295–312.

Lubinski, D. (Ed.). (2004). *Cognitive abilities: 100 years after Spearman's (1904) "'General intelligence,' objectively determined and measured". Journal of Personality and Social Psychology*, *86*. (pp. 96–199) (Special Section).

Mancuso, J. M. (2009). Assessment and measurement of health literacy: An integrative review of the literature. *Nursing and Health Sciences*, *11*, 77–89.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf–Gc framework. In D. P. Flanagan, J. L. Genshaft, P. L. Harrison, D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151-119). New York, NY US Guilford Press.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric research. *Intelligence*, *37*, 1–10.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (pp. 13–104) (3rd ed.). New York American Council on Education.

Morrow, P. C. (1983). Concept redundancy in organizational research: The case of work commitment. *Academy of Management Review*, *8*, 486–500.

Mõttus, R., Johnson, W., Murray, C., Wolf, M. S., Starr, J. M., & Deary, I. J. (2013). Towards understanding the links between health literacy and physical health. *Health Psychology*. http://dx.doi.org/10.1037/a0031439 (Advance online publication).

Murray, C., Johnson, W., Wolf, M. S., & Deary, I. J. (2011). The association between cognitive ability across the lifespan and health literacy in old age: The Lothian Birth Cohort 1936. *Intelligence*, *39*, 178–187.

Neddenriep, C., Hale, A., Skinner, C., Hawkins, R., & Winn, B. (2007). A preliminary investigation of the concurrent validity of reading comprehension rate: A direct, dynamic measure of reading comprehension. *Psychology in the Schools*, *44*, 373–388.

Nielson-Bohlman, L., Panzer, A. M., & Kindig, D. A. (Eds.). (2004). *Health literacy; A prescription to end confusion.* Institute of Medicine, Washington, D.C. The National Academies Press.

Page, E. B. (1986). The disturbing case of the Milwaukee Project. In H. H. Spitz (Ed.), *The raising of intelligence: A selected history of attempts to raise retarded intelligence* (pp. 115–140). Hillsdale, NJ Lawrence Erlbaum Associates.

Reeve, C. L., & Basalik, D. (2010). Average state IQ, state wealth and racial composition as predictors of state health statistics: Partial support for 'g' as a fundamental cause of health disparities. *Intelligence*, *38*, 282–289.

Reeve, C. L., & Bonaccio, S. (2011a). Nature and structure of intelligence. In T. Chamorro-Premuzic, A. Furnham, & S. von Stumm (Eds.), *Handbook of individual differences* (pp. 187–216). Oxford, UK Wiley-Blackwell.

Reeve, C. L., & Bonaccio, S. (2011b). On the myth and the reality of the temporal validity degradation of general mental ability test scores. *Intelligence*, *39*, 255–272.

Reeve, C. L., & Charles, J. E. (2008). Survey of opinions on the primacy of g and social consequences of ability testing: A comparison of expert and non-expert views. *Intelligence*, *36*, 681–688.

Reeve, C. L., & Lam, H. (2005). The psychometric paradox of practice effects due to retesting: Measurement invariance and stable ability estimates in the face of observed score changes. *Intelligence*, *33*, 535–549.

Reza, N., Hassan, M., Hayedeh, N., Ramin, H., & Elham, S. (2012). Original research: Teach back and pictorial image educational strategies on knowledge about diabetes and medication/dietary adherence among low health literate patients with type 2 diabetes. *Primary Care Diabetes*. http://dx.doi.org/10.1016/j.pcd.2012.11.001.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis*, *Vol. 2*, New York, McGraw-Hill.

Ruch, W. W., Stang, S. W., McKillip, R. H., & Dye, M. (2001). *Employee Aptitude Survey Technical Manual* (2nd ed.) Glendale, CA Psychological Services, Inc.

Scarr, S. (1993). Biological and cultural diversity: The legacy of Darwin for development. *Child Development*, *64*, 1333–1353.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.

Schwab, D. E. (1980). Construct validity in organizational behavior. In B. M. Staw, & L. L. Cummings (Eds.), *Research in organizational behavior*, *Vol. 2*. (pp. 3–43). Greenwich, CT JAI Press.

Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist*, *42*, 137.

Spearman, C. (1904). "General intelligence", objectively determined and measured. *American Journal of Psychology*, *15*, 201–292.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York MacMillan.

Spitz, H. H. (1986). *The raising of intelligence: A selected history of attempts to raise retarded intelligence.* Hillsdale, NJ Erlbaum.

Sternberg, R. J., Totff, B., & Grigorenko, E. L. (2009). Teaching triarchically improves school achievement. In R. J. Sternberg, J. C. Kaufman, & E. L. Grigorenko (Eds.), *The essential Sternberg: Essays on intelligence, psychology, and education* (pp. 195–215). New York, NY, US Springer Publishing.

Toulmin, S. (1961). *Foresight and understanding: An enquiry into the aims of science.* Bloomington Indiana University Press.

Weiss, B. D., Mays, M. Z., Martz, W., Castro, K. M., DeWalt, D. A., Pignone, M. P., et al. (2005). Quick assessment of literacy in primary care: The newest vital sign. *Annals of Family Medicine*, *3*, 514–522.

Williams, W. M., Blythe, T., White, N., Li, J., Gardner, H., & Sternberg, R. J. (2002). Practical intelligence for school: Developing metacognitive sources of achievement in adolescence. *Developmental Review*, *22*, 162–210.

Wolf, M. S., Curtis, L. M., Wilson, E. A. H., Revelle, W., Waite, K. R., Smith, S. G., et al. (2012). Literacy, cognitive function, and health: results of the LitCog study. *Journal of General Internal Medicine*, *27*, 1300–1307.

World Health Organization (WHO) (1998). *Health promotion glossary.* Division of Health Promotion, Education and Communications Health Education and Health Promotion Unit. Geneva World Health Organization.