Intelligence 36 (2008) 681–688

Contents lists available at ScienceDirect

Intelligence

Survey of opinions on the primacy of *g* and social consequences of ability testing: A comparison of expert and non-expert views ☆

Charlie L. Reeve *, Jennifer E. Charles

University of North Carolina Charlotte, United States

ARTICLE INFO

Article history:
Received 7 March 2008
Received in revised form 27 March 2008
Accepted 28 March 2008
Available online 2 May 2008

Keywords:
g
Cognitive ability
Intelligence
Ability testing
Expert opinion

ABSTRACT

The current study examines the views of experts in the science of mental abilities about the primacy and uniqueness of *g* and the social implications of ability testing, and compares their responses to the views of a group of non-expert psychologists. Results indicate expert consensus that *g* is an important, non-trivial determinant (or at least predictor) of important real world outcomes for which there is no substitute, and that tests of *g* are valid and generally free from racial bias. Experts did not reach consensus on issues such as the degree to which specific abilities or combinations of non-cognitive traits can yield predictive validities comparable to that of *g* alone, the predictive validity of *g* for non-technical work outcomes (e.g., contextual performance), and the nature and implications of race differences in intelligence. Second, a comparison of responses from experts and a group of applied psychologists reveals several discrepant beliefs between these groups, primarily dealing with the primacy of *g*, susceptibility of ability tests to racial bias, and the potential value of ability testing. Results are discussed in terms of directions for future research and shared responsibility for various groups of researchers to enhance dissemination of research to relevant audiences.

1. Introduction

Following the influence of his cousin, Charles Darwin, Francis Galton (1869) pioneered the study of individual differences in intelligence. Although Galton lacked the necessary statistical tools to investigate the structure of mental abilities, his writings indicate that he conceived of a general mental ability as the primary characteristic differentiating individuals in intellectual endeavors. Since that time, the science of mental abilities has arguably been one of the most heavily researched, and perhaps most controversial, topics in psychology. Indeed, there has been an active debate surrounding the primacy or uniqueness of general cognitive ability (i.e., 'g'; e.g., Gardner, 1993; Jensen, 1998; Spearman, 1904; Sternberg, 1985; Thurstone, 1947), its genetic, biological and environmental precursors (Plomin & Spinath, 2004), its role in or impact on academic, occupational, and social outcomes (e.g., Gordon, 1997; Gottfredson, 1997a; Kuncel, Hezlett, & Ones, 2004; Schmidt & Hunter, 1998; Sternberg & Hedlund, 2002; Vasquez & Jones, 2006), and its distribution across ethnic groups (e.g., Lynn, 1997; Ogbu, 2002; Rushton & Jensen, 2005). Likewise, there continues to be a disagreement about the perceived value, business necessity, and broader social consequences of cognitive ability testing in academic and employment settings (e.g., Herrnstein & Murray, 1994; Messick, 1995; Murphy, 2002; Outtz, 2002; Sackett, Schmitt & Ellingson, 2001; Schmidt & Hunter, 1998; Vasquez & Jones, 2006). That such disagreements still exists more than 130 years after Galton pioneered this field is evidenced by recent exchanges in academic journals, such as *Human Performance* (Viswesvaran & Ones, 2002) and *Perspectives on Psychological Science* (Brody, 2007; Gottfredson, 2007; Hunt & Carlson, 2007; Sternberg & Grigorenko, 2007), and the continuous publication of books taking various positions on these topics (e.g., Fish, 2001; Gould, 1996; Herrnstein &

☆ The authors wish to thank Douglas Detterman for granting access to the ISIR membership database, and Silvia Bonaccio for her helpful feedback on earlier drafts.

* Corresponding author. 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).

0160-2896/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2008.03.007

Murray, 1994; Jacoby & Glaubermann, 1995; Jensen, 1998; Murdoch, 2007).

To take but one example, consider the question of whether ability tests predict future performance in academic and occupational settings. On one hand, numerous authors have commented that the enormous volume of evidence supporting the predictive validity of general cognitive ability tests is unequivocal and beyond dispute (e.g. Brand, 1996; Gordon, 1997; Gottfredson, 2002; Jensen, 1998; Schmidt & Hunter, 1998). Schmidt (2002) for example states, "Given the overwhelming research evidence showing the strong link between general cognitive ability (GCA) and job performance, it is not logically possible … to have a serious debate over whether GCA is important for job performance" (p. 187). On the other hand, there continue to be assertions in highly visible outlets stating that there is little evidence that scores on such tests relate to real world success. For example, Vasquez and Jones (2006) recently stated "Standardized tests are thus not sufficiently predictive of future performance. Individuals are not necessarily more meritorious if they obtain the highest scores on standardized tests, thus rendering invalid the argument that students with the highest scores should have priority in admissions" (p. 138). At a minimum, such discrepant views certainly give the appearance of continuing controversy, even for what is arguably an empirical question.

At first glance, such discrepant views may give the impression that experts remain deeply divided over almost all aspects of the science of mental abilities. However, given the longevity and volume of research in the science of mental abilities, it is likely that there are areas of scientific consensus. We believe the appearance of controversy regarding a number of issues is driven by two factors. First, given the volume and increasing technical sophistication of the empirical literature, we admit it can be quite difficult for even research scientists to determine where scientific consensus has been achieved and which propositions and hypotheses are still legitimately in question. Second, the highly visible non-scientific commentaries (e.g., Gould, 1996; Murdoch, 2007) continue to give the impression that the field is in disarray. Indeed, such a sentiment was expressed by Reeve and Hakel (2002) who stated,

> "… scientific research on intelligence has often met with fierce public opposition. Even within the scientific community, the debate is often sidetracked by misunderstandings and misconceptions. The same questions are asked repeatedly, false claims and criticisms are based on misconstrued or misunderstood evidence, and important questions remain ignored. This wastes the resources, time, and energy of partisans, scientists, and the public." (p. 69).

As such, the purpose of this paper is twofold. First, we seek to determine where there is a consensus and where there is genuine controversy regarding the primacy and uniqueness of *g* and the social implications of ability testing among experts in the science of mental abilities. In this sense, we believe the results of such a survey can provide a valuable guide for future research in that we will be able to more efficiently and effectively focus on areas for which there is genuine controversy. Importantly, we do not propose that the results of such a survey will serve to resolve any genuine controversies about the science of mental abilities or its application. Indeed, as Snyderman and Rothman (1987) noted, "issues of fact are not settled via consensus." Rather, we posit that a survey of experts' opinions about ability testing is an effective way to insert a clear picture of informed opinion into both the scientific and public debates.

Second, we compare experts' opinions to those of a non-expert group of psychologists (namely, IO psychologists and practitioners as reflected by the results of Murphy, Cronin & Tam, 2003). We believe this second purpose has value in two respects. First, by examining the consistency of opinions and beliefs between experts and applied psychologists, we can gain some insight into how well the empirical literature in the science of mental abilities is being disseminated amongst a key consumer group. Second, the finding of discrepant beliefs would seem to indicate areas for which both groups need to enhance the cross-dissemination of relevant research to the other group. That is, we believe scholars in both groups have dual responsibilities: the creation of knowledge, and the effective dissemination of that knowledge to relevant audiences. We hope the results of this survey will serve to enhance both endeavors.

2. The use of surveys of scientific opinion

Surveying experts to discern areas of scientific consensus, and/or to contrast expert and non-expert views, has been used in several scientific disciplines. For example, such surveys have been conducted on a variety of topics including the ethics of human euthanasia (Simon, Kar, Hinz & Beck, 2007), autism (Gilliam & Coleman, 1982), epilepsy treatment (Karceski, Morrell, & Carpenter, 2001), social implications of nanotechnology (Besley, Kramer, & Priest, 2008), policy positions of political parties (Laver, 1998), the effects of biodiversity on ecosystems (Schläpfer, 1999), and climate change (Nordhaus, 1994).

With regard to the beliefs about cognitive abilities and cognitive ability testing, there have been several previous attempts to discern consensus among experts. In 1921, the editors of the *Journal of Educational Psychology* convened a meeting of 14 distinguished experts with the goal of finding a consensus definition of the broader term "intelligence." Sternberg and Detterman (1986) published the results of another symposium of 25 scholars. Again their primary focus was on the definition of the term "intelligence." Snyderman and Rothman (1987) used a survey technique to understand the beliefs and opinions on intelligence and intelligence testing among a broad population of social scientists.

Most recently, Murphy et al. (2003) published the results of a survey of industrial-organizational (IO) psychologists' beliefs about cognitive ability and cognitive ability testing. Using exemplary survey construction techniques based on a content sampling procedure, Murphy et al. developed a survey of 49 statements that reflected claims or assertions made in recent debates about cognitive abilities and cognitive ability testing. Their survey dealt with five general topics that reflected two underlying themes (primacy of *g* vs. societal concerns): (a) the importance or uniqueness of *g*, (b) the construct validity of ability tests, (c) the association between ability tests and job requirements, (d) value of alternatives or

additions to ability tests, and (e) societal impact of ability testing. Based on the results of their study, they concluded that, among a group of applied psychologists (primarily IO psychologists), "there was consensus over many items reflecting societal concerns and polarized opinions over many items reflecting beliefs about the primacy of *g*" (p. 667).

Although we believe their survey's questions represent a content valid sampling of the domain of issues surrounding general cognitive ability and the use of cognitive ability tests, we question the extent to which these results reflect the opinion of experts in the science of mental abilities. That is, while some IO psychologists certainly can be considered experts in the science of mental abilities, it seems unlikely that the population of IO psychologists *as a whole* would be considered experts in intelligence theory and measurement.[1] Given that very few IO graduate programs offer a course in differential psychology and that a course in psychometrics or measurement is not a standard requirement of many IO programs (Schmidt, 2002), it is unclear to what degree the opinions of that population of applied psychologists accurately reflect the state of the science as perceived by experts. Rather, we believe Murphy et al.'s (2003) results are more appropriately positioned as a benchmark by which to evaluate how well the science of mental abilities is being disseminated among applied psychologists.

3. Method

3.1. Participants and procedures

To obtain an expert sample, we first built a list of potential names from three sources. First, we included the 25 individuals currently on the editorial board of the journal *Intelligence*. Second, we included all registered members (as of April, 2007) of the International Society of Intelligence Researchers (ISIR). Third, to include people who were active contributors to the science mental abilities by may not have been captured by the first two sources, we included any person who had published three or more articles in *Intelligence* over the last 3 years (i.e., January 2004 and April 2007). This process resulted in 99 unique individuals. Active emails could not be obtained for 5 individuals, thus the sampling frame had an $N=94$.

We then sent emails to the 94 individuals asking them to participant in an anonymous web-based survey regarding beliefs about cognitive ability and cognitive ability testing. The content, item order, and response scale of the survey was identical to that used by Murphy et al. (2003). The 49 items are shown in our Table 1. Approximately one month later, we sent reminder emails to the same group. Three weeks later the web-based survey was taken off-line.

We obtained a response rate of 38.3% ($N=36$). Acknowledging that our initial sampling process may have included non-expert individuals (e.g., anyone with an interest in intelligence can join ISIR), we filtered the sample to include only individuals with a doctorate degree, and having at least five career publications on the topic of intelligence or testing. This resulted in a final expert sample of $N=30$. This sample size appears consistent with previous reports of expert opinions on intelligence (e.g., Gottfredson, 1997b; *Journal of Educational Psychology*, 1921; Sternberg & Detterman, 1986). Likewise, our response rate (38.3%) is similar to that obtained by Murphy et al. (2003; 31.5%).

While we acknowledge that our process to create a sampling frame may not have captured everyone who could arguably be considered an expert in the science of mental abilities, we believe it yielded a sample of respondents who qualify as experts. As noted, everyone in the sample has a doctorate degree. The average year of degree completion was 1981 (SD = 15.5 years) with a range from 1956 to 2006. Based on self-report data,[2] the average number of total publications was 116.6 (SD = 97.6; range = 5 to 475), and the average number of publications directly dealing with topics covered in the survey was 48.20 (SD = 57.79; range = 5 to 225). In this respect, the sample appears to be a knowledgeable set of active researchers. Most reported working in academic settings (87.6%), two worked in non-university research organizations, one individual worked for a test publisher, and one individual did not respond to this item. Most were male (90.0%) and all but one individual self-identified their race as White.

4. Results and discussion

Table 1 presents the descriptive statistics for each of the 49 items based on both the expert and applied psychologist data.[3] Given the orientation of the response scale, higher scores reflect the agreement side of the scale whereas lower scores reflect the disagreement side of the scale (scale values ranged from 1 = Strongly Disagree to 5 = Strongly Agree). Table 1 also includes the percentage of each group that agreed (i.e., endorsed agree or strongly agree) and disagreed (i.e., endorsed disagree or strongly disagree). The percentage of respondents indicating "no opinion" is not shown, but can be determined by subtracting the other two percentages from 100%. The item numbers shown in Table 1 reflect the order of items on the survey.

To facilitate comparison with the results of Murphy et al. (2003), we sorted the items into three categories (consensus items, polarized opinions, and neither consensus nor controversy) using the same criteria as those authors. "Consensus items" reflect those items for which less than 25% said no opinion, and the percentage of respondents saying agree (or disagree) was at least three times as large as the percentage saying disagree (or agree). Note, like Murphy et al., we are defining this level of agreement as "consensus," but we are not claiming unanimity of opinion. However, we agree with Murphy et al. that the assertion of consensus is reasonable when at least 75% of experts have an

[1] This comment is not meant to be a criticism of Murphy et al.'s (2003) sampling procedure. Indeed, those authors are explicit about purposely not restricting their sample to experts; rather, their purpose was to sample IO psychologists and practitioners broadly (see pp. 662-663).

[2] Although the high end of these ranges may seem large, the numbers are consistent with the publication records of a few individuals known to be within our sampling frame (e.g., Arthur Jensen).

[3] Please note that we use the terms "expert" and "applied psychologist" in this paper simply to distinguish between the two groups. As noted in the Introduction, this is in no way meant to imply that any given applied psychologist could not be an expert on this topic, or that experts on this topic do not do applied work.

Table 1
Survey results for the expert and applied psychologist groups

		Experts					Applied psychologists				
		M	SD	%D	%A	Type	M	SD	%D	%A	Type
Consensus items among experts											
3	GCA is measured reasonably well by standardized tests	4.47	.68	3.3	96.7	C	3.97	0.85	9.2	85.1	C
8	GCA will become increasingly important as the skills and knowledge required for good job performance become more complex	4.30	.88	6.7	86.7	C	3.97	0.92	11.3	81.0	C
11	GCA enhances performance in all domains of work	4.30	.53	0.0	96.7	C	3.43	1.12	29.7	62.0	N
29	Tests of non-cognitive traits are useful supplements to *g*-loaded tests …, but they cannot substitute for … *g*	4.27	.69	3.3	93.3	C	3.67	1.00	17.1	69.5	C
10	GCA is the most important trait determinant of job and training performance	4.10	.92	6.7	76.7	C	3.29	1.16	31.9	53.8	P
34	In jobs where CA is highly important, Blacks are likely to be underrepresented	4.03	.72	0.0	76.7	C	3.52	0.99	17.0	58.6	C
37	General cognitive ability tests are fair	4.03	1.16	10	76.7	C	3.62	0.99	14.2	62.7	C
2	GCA is the most important individual difference variable	4.00	1.02	13.3	76.7	C	3.08	1.18	41.5	44.7	P
26	The use of CATs in selection leads to more social justice than their abandonment	4.00	1.02	6.7	73.3	C	3.43	0.94	16.7	51.2	N
36	Professionally developed CATs are not biased against members of … minority groups	4.00	1.14	13.3	73.3	C	3.31	1.20	29.2	51.0	N
1	There is no substitute for GCA	3.97	1.27	20	70	C	3.47	1.22	30.2	62.4	P
42	Changes in test formats rarely change adverse impact without also changing the constructs being measured	3.97	.76	3.3	76.7	C	3.31	0.91	20.7	48.5	N
4	There is more to intelligence than what is measured by a standard cognitive ability test	3.87	1.11	16.7	76.7	C	4.16	0.80	5.6	89.0	C
17	The predictive validity of CATS depends on how performance criteria are defined and measured	3.80	1.03	16.7	80	C	4.17	0.78	6.0	90.0	C
28	CATs should almost always be a part of a personnel selection system	3.80	1.03	16.7	70	C	3.32	1.17	31.9	53.8	P
9	Different jobs are likely to require different types of cognitive abilities	3.77	1.19	23.3	73.3	C	3.94	0.91	11.9	82.6	C
33	The multidimensional nature of job performance necessitates the use of both cognitive and non-cognitive selection measures	3.77	1.10	20	76.7	C	4.33	0.77	3.9	90.9	C
45	Average scores on GCA tests are related to the effectiveness of an organization	3.77	.86	3.3	66.7	C	3.20	0.91	22.0	39.7	N
23	If two selection batteries show similar levels of CRV and similar costs, the battery that shows less AI should usually be preferred	3.70	.99	16.7	66.7	C	4.28	0.77	4.0	91.4	C
25	Employers should hire on the basis of the best predictors of performance, even if this leads to adverse impact	3.70	1.06	16.7	60	C	3.35	1.13	29.1	55.5	N
27	Tests of GCA can be used to create equal opportunities for all	3.67	.99	16.7	63.3	C	3.25	0.98	23.2	44.7	N
41	**Persons with lower levels of GCA will have to settle for lower level jobs**	3.67	.84	13.3	70.0	C	2.73	1.00	50.8	28.2	N
12	GCA is little more than academic intelligence	2.10	1.06	80	20	C	2.30	1.01	70.8	17.2	C
13	GCA tests measure constructs that are not required for successful job performance	1.90	1.09	76.7	13.3	C	2.31	1.04	68.1	19.5	C
22	Choosing to use CATs implies a willingness to accept the social consequences of racial discrimination	1.90	1.03	80	13.3	C	2.35	1.14	68.0	21.3	C
44	GCA tests show levels of validity too low to justify their negative social consequences	1.60	.67	90.0	0.0	C	2.13	0.93	74.0	10.1	C
14	The validity of CATs for real-life outcomes is low	1.57	.82	96.7	3.3	C	2.19	0.96	75.6	13.6	C
Polarized opinions among experts											
5	GCA accounts almost totally for the predictive validity of ability tests	3.40	1.33	33.3	63.3	P	3.05	1.08	39.6	44.1	P
30	**Combinations of specific aptitude tests have little advantage over measures of GCA in personnel selection**	3.27	1.20	33.3	53.3	P	0.66	1.01	56.9	25.1	N
24	If it is possible to develop a non-cognitive test battery with equal validity and less adverse impact than a cognitively loaded battery, the non-cognitive battery would be preferable	3.13	1.14	33.3	43.3	P	3.73	1.03	16.6	68.2	C
18	Two predictors that have equal criterion-related validity and equal costs will have equal value to an organization	3.10	1.06	36.7	43.3	P	2.40	1.17	65.9	22.6	C
6	Any combination of two or more specific aptitude tests is actually a measure of GCA	3.03	1.25	43.3	46.7	P	2.63	0.98	54.9	23.9	N
43	Massive societal changes will be necessary to significantly affect the adverse effects of CATs	3.00	1.11	36.7	33.3	P	3.49	1.06	22.5	51.6	N
39	Racial differences produced by CATs are substantially higher than racial differences on measures of job performance	2.93	1.23	40	36.7	P	3.16	1.02	29.6	43.1	N
21	Diversity in the workplace gives organizations competitive advantage	2.90	.92	33.3	30	P	3.74	0.87	8.3	67.6	C
15	Tacit knowledge is a form of practical intelligence, which explains aspects of performance not accounted for by … *g*	2.83	1.09	46.7	40	P	0.45	0.87	14.0	54.8	N
38	The use of GCA tests in selection typically results in a higher proportion of false negatives among minority groups	2.67	1.35	53.3	33.3	P	2.94	0.99	33.8	30.8	P
Neither consensus nor controversy among experts											
35	The belief that CATs are fair is most widespread among those groups who do not suffer the adverse effects of such tests	3.27	1.11	26.7	50	N	3.88	0.92	9.4	73.3	C
7	Unlike differences in complexity levels, differences between jobs in task make-up do not affect the validity of GCA tests	3.23	1.04	26.7	43.3	N	2.85	1.01	42.6	31.6	P
16	Although CATs are the best predictors of technical performance, they are not the best predictors of … contextual performance	3.20	.96	26.7	46.7	N	3.75	0.91	11.0	68.1	C
20	The dollar value of diversity can be measured	3.10	.80	20	33.3	N	2.93	1.05	37.1	33.6	P
31	Tacit knowledge contributes over and above g to the prediction of job performance	3.00	1.05	30	36.7	N	3.43	0.84	11.9	52.1	N
40	Blacks and Whites differ more on GCA than on virtually any other attribute	2.87	1.17	46.7	26.7	N	2.82	0.93	37.6	23.3	N
49	A workforce selected on the basis of actual performance would be less racially segregated than workforce selected on the basis of CATs	2.80	1.19	43.3	26.7	N	3.26	1.08	30.1	50.5	N
46	People are hired into organizations largely on the basis of their scores on CATs	2.77	.90	43.3	20	N	2.20	0.85	71.1	9.1	C
48	The choice between using CATs versus other non-cognitive selection measures is ultimately a question of values	2.67	1.21	50	23.3	N	2.79	1.20	52.6	36.0	P

Table 1 (*continued*)

		Experts					Applied psychologists				
		M	SD	%D	%A	Type	M	SD	%D	%A	Type
Neither consensus nor controversy among experts											
47	**There is a tradeoff between the cost-effective use of CATs and social responsibility in selection practices**	2.60	.93	56.7	23.3	N	3.33	1.02	26.0	54.8	N
19	**There are combinations of non-cognitive measures with CRV comparable to that achieved by CATs**	2.43	1.07	66.7	26.7	N	3.38	1.05	23.4	53.8	N
32	Score banding reduces the reliability of measurement but should be used to reduce disparate impact	2.23	.94	63.3	10	N	2.93	0.99	34.4	32.9	P

Note. Some item text shown here has been abbreviated or changed to acronyms for display purposes; the item text used in the actual surveys was verbatim from Murphy et al. (2003). GCA = general cognitive ability; CATs = cognitive ability tests; AI = adverse impact; CA = cognitive ability. Expert $N = 30$. Results for applied psychologists are those reported by Murphy et al. (2003). Items in bold font indicate that the majority of the comparison group responded opposite that of the expert group (e.g., majority of experts agreed with an item whereas a majority of the comparison group disagreed with the item). "Type" refers to the item classifications as explained in the text; C = "consensus," P = "polarization," and N = "neither consensus nor polarization."

opinion and there is a decidedly clear majority on one side of the issue. "Polarized opinions" reflect those items for which more than 30% said agree and more than 30% said disagree. We believe the classification of "polarized" is reasonable when more than 60% of experts have a clear opinion, but they tend to be split on opposite sides of the issue. Finally, items which did not meet the criteria for consensus or polarized were placed into a "neither consensus nor controversy" category.

Note that there are two columns in Table 1 indicating each item's classification; one based on the expert responses, and one based on the applied psychologist responses, respectively. These columns can be used to quickly compare the groups on each item. However, given the first purpose of the paper is to discern expert views, the items are sorted in Table 1 according to the expert responses. Finally, the items for which the majority of applied psychologists responded in a manner opposite that of the expert group are shown in bold font text (e.g., majority of experts agreed with an item whereas a majority of the comparison group disagreed with the item).

4.1. Areas of consensus among experts

Consensus was reached on 27 of the 49 items on the survey. Most of the items for which there was consensus dealt with issues concerning the existence and interpretation of empirical evidence. Specifically, the results suggest there is consensus among experts that *g* is an important, non-trivial determinant (or at least predictor) of important real world outcomes, and that tests of *g* are valid, and generally free from racial bias. Additionally, there appears to be a consensus for the idea that, although there is certainly more to intelligence than just *g* and that performance is determined by more than *g*, there is no substitute for *g*. These results appear to be in line with meta-analytic reviews of the empirical literature (e.g., Kuncel et al., 2004; Schmidt & Hunter, 1998) as well as previous expert commentaries (e.g., Gottfredson, 1997b; Jensen, 1998; Lubinski, 2000; Schmidt, 2002). These results also clearly contradict what might be termed "anti-testing, non-evidenced based commentaries" (e.g., Gould, 1996; Murdoch, 2007; Vasquez & Jones, 2006).

In addition to the consensus view on those issues, we also found expert consensus on the items reflecting beliefs about the implications of testing. For example, there is expert consensus that tests of *g* can be used fairly for personnel decision-making, that their use enhances social justice, and that such tests can assist in creating equal opportunities for everyone. These results clearly suggest that experts hold positive beliefs about the current and potential social value of cognitive assessments.

4.2. Areas of non-consensus among experts (issues likely in need of more research)

Examination of the set of items classified as "polarized" or "neither consensus nor polarization" provides insight into the issues for which additional research would be productive. We do not comment on each specific item, but rather discuss the apparent themes.

First, there appears to be a disagreement among the experts regarding the degree to which specific abilities, tacit knowledge, or combinations of non-cognitive traits, can yield predictive validities comparable to that of *g* alone, or add meaningful amounts of incremental validity over *g*. Several items presented variations on this theme (e.g., items 15, 24, 30, 31). This suggests a critical area in need of additional research, or at the least, effective quantitative summaries of the rather sizeable extant literature on this issue. Given that there are also differences in views on this issue relative to applied psychologists, we say more on this issue below.

Second, whereas there is clear consensus that *g* predicts what would be considered technical or core aspects of academic and job performance, there appears to be uncertainty about the degree to which *g* predicts "contextual performance" or other work adjustment outcomes. Contextual performance generally refers to those aspects of employee behavior that are discretionary and not explicitly recognized as a formal part of the job, but which contribute to and promote the effective functioning of the organization (Organ, Podsakoff, & MacKenzie, 2006). The lack of consensus in this area suggests a fruitful avenue for research that would further clarify the nature of the *g*-nexus.

Third, it would appear that questions regarding the nature of race differences in intelligence, and the implications of adverse impact, are still in need of additional research. It should be kept in mind that investigating views on the nature of race differences was not the focus of the survey; as such, there were only a few general items relating to this issue. Clearly this issue is complex and multifaceted; we caution readers from making strong inferences on the basis of these few items. Nonetheless, a global evaluation of the results does suggest a few general trends. For example, there appears to be some consensus (but not unanimity) among experts that professionally developed tests are not biased against minority groups. At

the same time, there are clearly some unresolved issues. Two items in particular (items 39 and 43) which deal with the nature of racial differences reveal polarized opinions. Thus, although this will undoubtedly continue to be a highly politicized and polarized line of inquiry, these results suggest that additional research is needed to better understand this phenomenon.

4.3. Comparison of experts and applied psychologists (implications for research and dissemination)

Our second purpose was to compare experts' opinions and beliefs about cognitive abilities and ability testing to those of applied psychologists so that we might gain some insight into how effectively this literature is being disseminated among applied psychologists. We start by noting the primary areas where experts and applied psychologists hold similar views. The results indicate both groups agree that *g* is an important, non-trivial construct that influences job performance, that its importance increases as the complexity of the job increases, and that tests of *g* are construct valid and show useful levels of criterion-related validity with respect to technical aspects of job performance. There also appears to be agreement that *g* is necessary but not sufficient for performance in most domains. That is, the multifaceted nature of job performance, requires additional abilities, skills, and non-cognitive traits be assessed for a more complete understanding of the precursors of work performance. If anything, we believe this finding essentially validates the survey; we know of no scientific psychologist to have ever claimed performance is solely determined by a single variable, *g* or otherwise.

One of the most apparent areas of disagreement between experts and applied psychologist concerns the primacy of *g*. Though applied psychologists appear to agree with the expert group that *g* is an important individual difference variable, there seems to be disagreement between experts and applied psychologists in terms whether *g* is the "most important individual difference variable" and whether there is a substitute for it. In particular, in contrast to experts, applied psychologists failed to reach consensus on items stating that *g* enhances performance in all domains of work (item 11), that *g* is the most important trait determinant of job and training performance (item 10), that *g* is the *most* important individual difference variable (item 2), and that there is no substitute for *g* (item 1). For example, whereas almost 30% of applied psychologists disagreed with the statement that *g* enhanced performance in all domains of work, none of the experts disagreed and 96.7% agree with that statement. Likewise, whereas more than 30% of applied psychologists disagreed that *g* is the most important trait determinant of performance, only 6.7% of experts disagreed and more than 75% agreed with the statement. This difference is most obvious on the item stating that *g* is the most important individual difference variable; more than 75% of experts agreed with the statement but applied psychologists were split evenly on this item with roughly 40% falling on opposite sides.

To some degree, it is possible that the discrepancy in views on this issue is a function of some ambiguity in the phrase "most important." It is possible that more applied psychologists than experts attach a non-technical meaning to such a phrase and as such are less willing to endorse a statement concerning importance. But why applied psychologists would do this, and experts as a rule would not, is unclear. Rather, that there is a clear consensus on this issue among experts, and disagreement on this issue among the applied psychologists, strikes us as an indicator of limited knowledge dissemination rather than an indicator of genuine scientific controversy. That is, if we define "primary determinant" as the strongest predictor (i.e., uniquely accounts for the greatest percentage of variance) of domain-specific performance differences, then whether *g* is the most important trait determinant of performance is an empirical question with a clear answer. Multiple large-scale meta-analyses (e.g., Hunter & Hunter, 1984; Kuncel et al., 2004; Salgado, Anderson, Moscoso, Bertua, & De Fruyt, 2003; Schmidt & Hunter, 1998) clearly demonstrate that measures of *g* consistently demonstrate the largest criterion-related validities of any trait measure. That IO psychologists on average appear to be unaware of this literature would seem to validate Frank Schmidt's (2002, p. 200) admonishment:

> "It would be irresponsible to ignore … the role of GCA in job performance. Yet, this does sometimes happen, even when the individual differences research is conducted by I/O psychologists … Why does this happen? One reason is that most I/O graduate programs no longer include a course in differential psychology."

A second prominent issue for which there were discrepant views across the two groups is that of racial bias, and the attendant issues stemming from adverse impact. Although the lack of racial bias is generally an accepted fact among experts, almost half of applied psychologists did not agree with the item indicating that professionally developed ability tests are unbiased (i.e., item 36). That is, despite acknowledging that *g* is an important, non-trivial construct that influences job performance, significant numbers of applied psychologists continue to believe that ability tests suffer from racial bias. We find this discrepancy to be particularly interesting given that the issue of test bias should be of profound concern for applied psychologists. As Reeve and Hakel (2002, p. 60) noted, "I-O psychologists need to be informed regarding research on the nature of race differences. It would seem difficult to use intelligence tests ethically or advise others about policy regarding adverse impact without being knowledgeable about the nature of group and individual differences." Thus we would expect that the curriculum in any graduate program training applied psychologists would include this literature. Although claims of bias can still be found in the popular press, we know of no reliable evidence that professionally developed intelligence tests, or the factor scores derived from such tests, are subject to cultural or measurement test bias. Indeed, Jensen's (1980) tome on the issue still stands as the definitive account; 30 years of additional research has only confirmed his conclusions. As such, we doubt the discrepancy in views between experts and applied psychologists is due to the existence of a large empirical literature showing evidence for bias that is accessible to applied psychologists yet remains unknown to intelligence and testing experts. Rather, we interpret this finding as an indicator of the need for better dissemination of this literature among applied psychologists; in particular, we concur with Schmidt's admonishment of IO graduate programs for failing to offer courses in psychometrics and differential psychology.

Third, although there was not a consensus among experts regarding the question of whether specific abilities, or a collection

of non-cognitive traits, can yield predictive validities comparable to *g* alone, it is interesting to note that the majority of applied psychologists hold views opposite the majority of experts. About two-thirds of the expert group disagreed with the statement that combinations of non-cognitive measures yield criterion-related validities comparable to *g*, but more than half (54%) of the applied psychologists believe this to be true. Similarly, a majority of experts (53%) agreed with the statement that combinations of specific abilities have little predictive advantage over measures of *g*, but over half of the applied psychologists (57%) disagreed. While these results certainly suggest additional research on such questions is likely to be fruitful, it is possible some of the apparent disagreement here stems from differences in the interpretation of the question. For example, with regard to job performance criteria, there is substantial evidence from industrial psychology to suggest that, across the broad spectrum of jobs in the U.S. economy, *g* predicts performance better than any other individual specific attribute, and predicts about as well as optimally weighted composites from batteries of cognitive tests (e.g., Gottfredson, 1997a; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Ree, Earles, & Teachout, 1994). However, the work by Lubinski and Benbow (e.g., Lubinski, Benbow, Webb, & Bleske-Rechek, 2006; Park, Lubinski, & Benbow, 2007) clearly demonstrates that ability-tilt (i.e., profile of specific abilities) contributes meaningfully to the prediction of the domain in which people are likely to succeed. The resolution of the apparent paradox in these two sets of findings lies in the details of the precise question asked. When predicting inter-individual differences in a general criterion, it seems to be true that specific abilities rarely yield large increments in validity. However, when examining questions of intra-individual variation, profiles of specific abilities appear to be highly informative. That is, general ability appears to account for most of the variation in the general level of achievement, and specific abilities appear to predict in which domain the achievement occurs.

Finally, the results also suggest a potential difference in values or perspectives between the expert group and applied psychologists regarding the potential use of tests. For example, there was a consensus among experts that employers should hire on the basis of the best predictors of performance, even if this leads to adverse impact; applied psychologists were more polarized on this item and reported a lower average agreement. Similarly, almost two-thirds of the experts agreed that tests of *g* can be used effectively to create equal opportunities, whereas less than half of the applied psychologists agreed with the item. We believe differences between groups on these types of items likely indicate a difference in values or perspectives rather than differences in knowledge of the empirical literature. The very nature of these questions seems to be less amenable to strictly empirical answers (though, this is not to say that empirical evidence cannot be useful in forming an opinion on these issues). Thus, these differences likely suggest areas in need of more thoughtful discussion between testing experts and applied groups, but it should be acknowledged that these are largely questions of values and social perspectives.

5. Conclusion

Though some commentaries give the impression of controversy regarding the importance of cognitive abilities and the validity of ability testing, the results of this survey clearly demonstrate that there are areas of resounding consensus among experts. Our results indicate that there is consensus among experts in the science of mental abilities that *g* is an important, non-trivial determinant (or at least predictor) of important real world outcomes for which there is no substitute, and that tests of *g* are valid and generally free from racial bias. The areas for which we found evidence of continued controversy appear to deal with what might be considered more detailed questions rather than core or fundamental questions. For instance, there appears to be lack of consensus regarding the degree to which specific abilities contribute meaningful variance above *g*, and as to the exact breadth of the *g*-nexus. These are important issues to be sure, but they are not the type of issues that call into question the fundamental importance of cognitive ability, or the validity and utility of ability tests in general.

Additionally, we found several areas for which non-expert psychologists held views dissimilar to experts. We believe it is important to better understanding why other groups of psychologists hold views contrary to intelligence experts. Such discrepancies may call into question the degree to which applied psychologists are being trained in important areas (as suggested by Schmidt, 2002), or the degree to which experts are disseminating their research to important audiences. In this sense, we encourage both experts and applied psychologists to consider ways to enhance the cross-boundary dissemination of their knowledge (e.g., editors of journals can develop special issues or consider non-traditional submissions; graduate programs could revisit their core curricula). However, it should be acknowledged that the science of mental abilities spans the basic-applied division; many of the questions found on the current survey clearly deal with applied issues of testing. Thus, differences in views between the expert group and the applied group may indicate areas where both groups need to accept a shared responsibility for enhancing the dissemination of research findings outside of their respective domains. Finally, we concur with a reviewer who noted that a legitimate debate entails two clear positions, a free exchange of arguments, weighing the evidence in favor and against the two positions, and reaching informed conclusions. In this sense, we hope our results can help both experts and non-experts engage in healthy, intellectually honest debates on these issues.

References

Besley, J. C., Kramer, V., & Priest, S. (2008). Expert opinion on nanotechnology: risks, benefits, and regulation. *Journal of Nanoparticle Research*, *10*, 549–558.

Brand, C. (1996). *The g factor: General intelligence and its implications.* New York: Wiley.

Brody, N. (2007). Barriers to understanding racial differences in intelligence. *Perspectives on Psychological Science*, *2*, 214–215.

Fish, J. M. (Ed.). (2001). *Race and intelligence: Separating science from myth* Mahwah, NJ: Lawrence Erlbaum Associates.

Galton, F. (1869). *Hereditary genius: An enquiry into its laws and consequences.* London: Collins.

Gardner, H. (1993). *Multiple intelligences: The theory in practice.* New York: Basic Books.

Gilliam, J. E., & Coleman, M. (1982). A survey of knowledge about autism among experts and caregivers. *Behavioral Disorders*, *7*, 189–196.

Gordon, R. A. (1997). Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence*, *24*, 203–320.

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

Gottfredson, L. S. (1997). Editorial: Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence*, *24*, 13–23.
Gottfredson, L. S. (2002). Where and why *g* matters: Not a mystery. *Human Performance*, *15*, 25–46.
Gottfredson, L. S. (2007). Applying double standards to 'divisive' ideas. *Perspectives on Psychological Science*, *2*, 216–220.
Gould, S. J. (1996). *The mismeasure of man.* New York: Norton.
Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.
Hunt, E., & Carlson, J. (2007). Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science*, *2*, 194–213.
Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72–98.
Jacoby, R., & Glaubermann, N. (Eds.). (1995). *The bell curve debate: History, documents, opinions* New York, NY: Random House.
Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.
Jensen, A. R. (1998). *The g factor.* Westport, CT: Praeger.
Karceski, S., Morrell, M., & Carpenter, D. (2001). The Expert Consensus Guideline Series: Treatment of epilepsy. *Epilepsy & Behavior*, *2*, 1–50.
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.
Laver, M. (1998). Party Policy in Britain 1997: Results from an expert survey. *Political Studies*, *46*, 336–347.
Lubinski, D. (2000). Scientific and social significance of assessing individual differences: "Sinking shafts at a few critical points". *Annual Review of Psychology*, *51*, 405–444.
Lubinski, D., Benbow, C. P., Webb, R. M., & Bleske-Rechek, A. (2006). Tracking exceptional human capital over two decades. *Psychological Science*, *17*, 194–199.
Lynn, R. (1997). Geographical variation in intelligence. In N. Helmuth (Ed.), *The scientific study of human nature: Tribute to Hans J. Eysenck at eighty* Oxford, England: Pergamon.
McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationships between predictor and criterion domains. *Personnel Psychology*, *43*, 335–354.
Messick, S. (1995). Validity of psychological assessment: Validation of inferences from person's responses and performance as scientific inquiry into score meaning. *American Psychologist*, *50*, 741–749.
Murdoch, S. (2007). *IQ: A smart history of a failed idea.* Hoboken, NJ: John Wiley & Sons.
Murphy, K. (2002). Can conflicting perspectives on the role of "g" in personnel selection be resolved? *Human Performance*, *15*, 173–184.
Murphy, K. R., Cronin, B. E., & Tam, A. P. (2003). Controversy and consensus regarding the use of cognitive ability testing in organizations. *Journal of Applied Psychology*, *88*, 660–671.
Nordhaus, W. D. (1994). Expert opinion on climatic change. *American Scientist*, *82*, 45–51.
Ogbu, J. (2002). Cultural amplifiers of intelligence: IQ and minority status in cross-cultural perspective. In J. M. Fish (Ed.), *Race and intelligence: Separating science and myth* (pp. 241–280). Mahwah, NJ: Erlbaum.
Organ, D. W., Podsakoff, P. M., & MacKenzie, S. B. (2006). *Organizational citizenship behavior.* Thousand Oaks, CA: SAGE Publications.
Outtz, J. (2002). The role of cognitive ability tests in employment selection. *Human Performance*, *15*, 161–172.
Park, G., Lubinski, D., & Benbow, C. P. (2007). Contrasting intellectual patterns predict creativity in the arts and sciences: Tracking intellectual precocious youth over 25 years. *Psychological Science*, *18*, 948–952.
Plomin, R., & Spinath, F. M. (2004). Intelligence: Genetics, genes, and genomics. *Journal of Personality and Social Psychology*, *86*, 112–129.
Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g*. *Journal of Applied Psychology*, *79*, 518–524.
Reeve, C. L., & Hakel, M. D. (2002). Asking the right questions about *g*. *Human Performance*, *15*, 47–74.
Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, *11*, 235–294.
Sackett, P., Schmitt, N., & Ellingson, J. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, *56*, 302–318.
Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, *56*, 573–605.
Schläpfer, F. (1999). Expert estimates about effects of biodiversity on ecosystem processes and services. *Oikos*, *84*, 346–352.
Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, *15*, 187–210.
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.
Simon, A., Kar, M., Hinz, J., & Beck, D. (2007). *Attitudes toward terminal sedation: An empirical survey among experts in the field of medical ethics.* BMC Palliative Care 6:4. Retrieved December 11, 2007, from http://www.biomedcentral.com/1472-684X/6/4
Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist*, *42*, 137–144.
Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, *15*, 201–292.
Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York: Cambridge University Press.
Sternberg, R., & Detterman, D. K. (1986). *What is intelligence?* Norwood, NJ: Ablex Publishing.
Sternberg, R. J., & Grigorenko, E. L. (2007). The difficulty of escaping preconceptions in writing an article about the difficulty of escaping preconceptions: Commentary on Hunt and Carlson (2007). *Perspectives on Psychological Science*, *2*, 221–223.
Sternberg, R. J., & Hedlund, J. (2002). Practical intelligence, *g*, and work psychology. *Human Performance*, *15*, 143–160.
Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of. The Vectors of the Mind.* Chicago: University of Chicago Press.
Vasquez, M. J. T., & Jones, J. M. (2006). Increasing the number of psychologists of color: Public policy issues for affirmative diversity. *American Psychologist*, *61*, 132–142.
Viswesvaran, C., & Ones, D. S. (2002). Agreements and disagreements on the role of general mental ability (GMA) in industrial, work, and organizational psychology. *Human Performance*, *15*, 211–231.