The international personality item pool and the future of public-domain personality measures ☆

Lewis R. Goldberg [a], John A. Johnson [b,*], Herbert W. Eber [c], Robert Hogan [d], Michael C. Ashton [e], C. Robert Cloninger [f], Harrison G. Gough [g]

[a] *Oregon Research Institute, USA*
[b] *Department of Psychology, Pennsylvania State University, College Place, DuBois, PA 15801, USA*
[c] *Psychological Resources, USA*
[d] *Hogan Assessment Systems, USA*
[e] *Brock University, USA*
[f] *Washington University, St. Louis, USA*
[g] *University of California, Berkeley, USA*

Available online 25 October 2005

Abstract

Seven experts on personality measurement here discuss the viability of public-domain personality measures, focusing on the International Personality Item Pool (IPIP) as a prototype. Since its inception in 1996, the use of items and scales from the IPIP has increased dramatically. Items from the IPIP have been translated from English into more than 25 other languages. Currently over 80 publications using IPIP scales are listed at the IPIP Web site (http://ipip.ori.org), and the rate of IPIP-related publications has been increasing rapidly. The growing popularity of the IPIP can be attributed to five factors: (1) It is cost free; (2) its items can be obtained instantaneously via the Internet; (3) it includes over 2000 items, all easily available for inspection; (4) scoring keys for IPIP scales are

provided; and (5) its items can be presented in any order, interspersed with other items, reworded, translated into other languages, and administered on the World Wide Web without asking permission of anyone. The unrestricted availability of the IPIP raises concerns about possible misuse by unqualified persons, and the freedom of researchers to use the IPIP in idiosyncratic ways raises the possibility of fragmentation rather than scientific unification in personality research.

1. Introduction

This article continues a discussion about the viability of public-domain personality measures that began at the eighth European Conference on Personality (ECP) in 1996 (Costa & McCrae, 1999; Goldberg, 1999a, 1999b). At that 1996 conference, a new public-domain personality resource, the International Personality Item Pool, abbreviated IPIP (Goldberg, 1999a), was first introduced, amid substantial qualms about its potential usefulness (Costa & McCrae, 1999). Now, nearly 10 years later, we here consider (a) the degree to which the IPIP has lived up to its promise, (b) future directions that might be taken, and (c) more generally, whether public-domain resources such as the IPIP will in fact become a viable alternative to commercial personality inventories.

The article is divided into two major sections. The first section is descriptive and the second analytic. The first section describes the history and rationale behind the construction of the IPIP and then provides a detailed portrait of what the IPIP looks like today as a result of those efforts. The second section of the article analyzes the strengths and weaknesses of the IPIP based on feedback from users.

2. The IPIP, yesterday and today

2.1. The development of the international personality item pool

The stimulus behind the IPIP was a perception that "the science of personality assessment has progressed at a dismally slow pace since the first personality inventories were developed over 75 years ago" (Goldberg, 1999a, p. 7). In a previous historical review, Goldberg (1995) had discussed a number of causes for the field's lack of consensus on a scientifically reasonable taxonomic structure for human personality traits. One major cause was the negative *Zeitgeist* for personality research engendered by the writings of Mischel and the situation-personality debate (see also Johnson, 1999a, 1999b). In regard to personality-trait measurement, however, Goldberg (1999a) attributed the seeming lack of progress in part to the policies and practices of commercial inventory publishers who could regard certain scientific activities as detrimental to their pecuniary interests (an issue also raised by Eber, 2005). There are at least four kinds of scientific activities that are disallowed or discouraged by test publishers.

First, commercial publishers rarely permit, much less encourage, researchers to extract and present only portions of an inventory, change the item order, reword items, collate the items with items from other measures, or engage in similar kinds of experimentation. Most publishers see such activities as potential threats to the integrity of their instruments, and

therefore they typically require researchers to purchase and use an entire inventory "as is." Although some publishers grant licenses for special uses, investigators must request and pay for all such licenses to use only selected scales or items.

Second, most publishers disallow the use of their copyrighted inventories on the World Wide Web. Personality researchers have recently discovered that collecting and scoring personality responses via the Web is vastly more efficient than doing these things by hand, and that Web-based research allows one to tap into much larger and more diverse participant pools (Buchanan, Johnson, & Goldberg, 2005; Gosling, Vazire, Srivastava, & John, 2004). However, from the perspective of a publisher, a copy of an inventory that can be used by all research participants on the Web eliminates the need to buy an inventory booklet (or pay a fee for an electronic administration) for each participant in a study. Illegal photocopying of paper inventories has always been a source of lost revenue for test publishers, but a copy of an inventory posted on the Web could be down-loaded by anyone in the world. Most publishers, therefore, insist that researchers refrain from posting copyrighted inventories on the Web, denying researchers the advantage of this burgeoning new method for data collection.

Third, publishers differ in their practices on providing scoring keys for the scales in their inventories: Some keys are published in test manuals; some are available to researchers on request; and a few are not released at all. Publishers of commercial inventories who keep their scoring keys secret may do so for several reasons. First, it allows them to collect a fee for scoring each inventory. In addition, for instruments to be used in personnel selection, guarding the keys could prevent applicants from "cheating" on the inventory. However, without access to the scoring keys personality researchers will have difficulty conducting item-level analyses such as assessing the internal consistency of the scales in their participant samples.

Finally, Goldberg (1999a) has suggested that the vested interests of each publisher discourage continual efforts at further test development and discourage comparative-validity studies. Instead, inventory authors and publishers appear to seek a loyal following of dedicated users. This encourages insulated efforts designed to market rather than to improve their tests. Typically, an inventory manual provides tables of correlations between scale scores and various criteria as a claim for the validity and utility of the inventory. Seldom, however, are these correlations used to develop new scales or to improve the quality of existing scales. Although different researchers often use different measures in multivariate prediction studies, one rarely sees true comparative-validity studies (e.g., Ashton & Goldberg, 1973; Goldberg, 1972; Johnson, 2000) that pit two or more broad-band inventories against each other to see which provides the strongest pattern of predictions. Consequently, results from these insulated research programs become incommensurable, leading to the fragmentation rather than integration of knowledge (Johnson & Ostendorf, 1993).

2.2. The IPIP as an antidote for slow research progress

Goldberg (1999a) suggested that placing a set of personality items in the public domain might free researchers from the constraints imposed by copyrighted personality inventories. Hence, the International Personality Item Pool (IPIP) was born. Not only may researchers freely use the items in the IPIP in any way they see fit, but they can also rapidly and easily access the items from the IPIP Web site at http://ipip.ori.org/. Goldberg envisioned the IPIP Web site as a collaboratory, defined as "a computer-supported system that

allows scientists to work with each other, facilities, and data bases without regard to geographical location" (Finholt & Olson, 1997, p. 28). The IPIP Web site is intended to provide rapid access to measures of individual differences, all in the public domain, to be developed conjointly among scientists worldwide.

The IPIP has grown from an initial set of 1252 items to well over 2000 items, and new sets of items are added each year. Roughly 750 of the IPIP items had their origins in Dutch, in a project initiated by A.A. Jolijn Hendriks, Willem K.B. Hofstee, and Boele de Raad at the University of Groningen (Hendriks, 1997). Since the IPIP's inception, portions of the item pool have been translated, or are in the process of being translated, into Arabic, Bulgarian, Chinese, Croatian, Danish, Dutch, Estonian, Finnish, French, German, Hebrew, Hmong, Hungarian, Italian, Korean, Latvian, Norwegian, Persian, Polish, Romanian, Russian, Serbian, Slovene, Spanish, Swedish, Turkish, Vietnamese, and Welsh. A page at the IPIP Web site keeps the research community informed about such translation projects (including multiple projects in Chinese, German, Spanish, and Swedish), with e-mail links to the investigators involved.

The common format for IPIP items follows the recommendations of the Groningen research team (Hendriks, 1997), who suggested that single trait adjectives are often too abstract to allow one-to-one translations even in languages as linguistically close as English, Dutch, and German. Single adjectives without an explicit context are also more likely than contextualized phrases to generate different interpretations among different respondents. Furthermore, single adjectives cannot convey the complex nuances of personality description as well as phrases. Therefore, the format chosen for IPIP items is a short verbal phrase—more contextualized than a single trait adjective, but more compact than items found in many modern personality inventories. Whether these phases contain enough context is a matter of debate (Cloninger, 2004, 2005). Some examples of IPIP items include: Am able to disregard rules; Believe in an eye for an eye; Can read people like a book; Dislike being the center of attention; Enjoy the beauty of nature; Forget appointments; Get upset easily; Have gotten better with age.

The IPIP collaboratory is intended as an international effort to develop and continually refine a set of personality scales, all of which remain in the public domain, to be available for both scientific and commercial purposes. The rationale behind the IPIP collaboratory is that no one investigator alone has access to many diverse criterion settings, but the international scientific community has such access, and by pooling their findings the community will see faster progress in the science of personality measurement (Goldberg, 2005a).

The IPIP Web site currently includes three major types of information: (a) psychometric characteristics of the current set of IPIP scales, which are continuously being supplemented by new scales; (b) keys for scoring the current set of scales; and (c) the current total set of IPIP items, which is continuously being supplemented with new items. The web site also serves as a repository for reports of studies using IPIP items. In the future, the site may include raw data available for reanalysis and also serve as a forum for the discussion of psychometric ideas and research findings (Goldberg, 2005a).

2.3. IPIP scale-development procedures

Some IPIP scales have been designed to serve as proxies for the constructs measured by scales in major commercial inventories, thereby providing a public-domain alternative to these inventories. The scale-construction procedure used by Goldberg (1999a) is described

on the IPIP Web site as a process that combines empirical, rational-intuitive, and psychometric methods. The stages in that process are described below.

First, all available IPIP items are correlated with each of the original inventory scales, using a sample that has responded to both item pools. When the original scales are part of a multi-scale inventory, each IPIP item is categorized by the scale with which it has its highest correlation, and IPIP items are rank-ordered within each of the resulting categories by the size of those correlations. This ensures that all of the IPIP items selected for an IPIP scale correlate more highly with its criterion scale than with any of the others.

Next, the K highest positively correlating IPIP items and the K highest negatively correlating IPIP items are selected for the preliminary scale, with K being 1/2 the number of items that are desired in the final scale (typically $5 + 5 = 10$). If the correlations with the original scale are substantially higher within the set correlating positively than are the correlations within the set correlating negatively, or vice versa, the criterion for equal numbers of positively and negatively keyed items is relaxed, by trying to balance equal keying direction with high strength of association.

In the third stage, the content of the IPIP items is examined, noting any item pairs that are essentially identical in content, and the lowest correlating item from such redundant pairs is omitted. If any item is omitted using this redundancy criterion, then another item from the set of most highly correlating IPIP items is added. At this stage, the content of all of the selected items is examined to see if they tell a coherent story. Any item that does not seem to mirror the major story-line is omitted, and a new item from the set of most highly correlating items is added.

The fourth stage involves a standard reliability analysis of the items in the scale. Any item whose addition to the scale lowers the coefficient-α reliability of the resulting scale is omitted, and another item from the set of most highly correlating items is appraised. This process is continued until α is as high as is reasonable, without sacrificing too much breadth of content. This last stage requires some ingenuity, and thus this is the stage where an exact algorithm would be difficult to formulate.

The resulting IPIP scales are always labeled "preliminary," based on the assumption that they should be able to be improved by the use of more sophisticated procedures, such as those based on Item Response Theory (IRT). In addition, as new IPIP items are added to the pool, some of them should prove to be more valid indicators of the constructs being measures. Members of the worldwide research community are invited to use IRT models and any other techniques to improve the quality of the preliminary IPIP scales.

2.4. Measures available at the IPIP web site

At the present time, approximately 300 scales constructed from IPIP items are available. IPIP proxies have been developed to measure the constructs in the following broad-bandwidth inventories: The NEO-PI-R (Costa & McCrae, 1992), 16 Personality Factor Questionnaire (16PF: Conn & Rieke, 1994), California Psychological Inventory (CPI: Gough & Bradley, 1996), Hogan Personality Inventory (HPI: Hogan & Hogan, 1992), Temperament and Character Inventory (TCI: Cloninger, 1994), Multidimensional Personality Questionnaire (MPQ: Tellegen, in press), Jackson Personality Inventory (JPI-R: Jackson, 1994), Six-Factor Personality Questionnaire (6FPQ: Jackson, Paunonen, & Tremblay, 2000), and HEXACO Personality Inventory (Lee & Ashton, 2004). Other multiple-construct measures with IPIP proxies include the lexical Big-Five factor structure (Goldberg, 1992), the lexical

Alternative 7 (Saucier, 1997), the 45 facets in the Abridged Big Five-dimensional Circumplex model (AB5C: Hofstee, de Raad, & Goldberg, 1992), components of Emotional Intelligence (Barchard, 2001), and the BIS/BAS Inhibition/Activation System (Carver & White, 1994). Tables comparing the psychometric characteristics of the original scales with the IPIP proxies are available for all of these inventories. In general, the coefficient-α reliabilities of the IPIP scales generally match or exceed the reliabilities of the original scales, and the IPIP scales correlate highly with their parent scales; consequently, as would be expected, when the corresponding pairs of scales are corrected for the reliabilities of their components, the corrected values are typically quite high.

Because some of the IPIP scales derived from different inventories are quite similar to each other, and thus share the same scale label, there are fewer constructs than scales. Included at the IPIP Web site is an alphabetical index of roughly 175 such constructs (e.g., Activity Level, Borderline Personality Disorder, Complexity, Empathy, Impulse Control, Impression-Management, Irrational Beliefs, Locus of Control, Self-Monitoring), many of which are measured by two or more IPIP scales.

2.5. What is not available at the IPIP web site

Although Goldberg readily provides investigators with raw data and statistical summaries from research participants who have completed IPIP measures, those data are not yet included directly at the IPIP Web site. More controversially, however, there are no norms available there. Goldberg's position is that one should be very wary of using "canned norms" because it is not obvious that one could ever find a population of which one's present sample is a representative subset. He has argued that most "norms" are misleading, and therefore they should not be used. IPIP consumers who wish to use norms are encouraged to develop local norms from their own samples.

The IPIP Web site remains agnostic about the development of "validity" indices for IPIP scales. Many inventory developers assume that such indices are necessary, but some research (e.g., Costa & McCrae, 1997; Johnson, 2005a; Kurtz & Parrish, 2001; Piedmont, McCrae, Riemann, & Angleitner, 2000) suggests otherwise. Included at the IPIP Web site is a description of methods for constructing validity indices for IPIP scales, including (a) overuse of the same response option, (b) response commonality versus deviancy, (c) semantic inconsistency, and (d) social-desirability biases. Under conditions of anonymity, the impact of allegedly invalid responding on the integrity of data appears to be negligible (e.g., Johnson, 2005a), but whether dissembling is a problem in non-anonymous evaluative conditions such as employee selection is another issue (Ones & Viswesvaran, 1998). However, although the IPIP Web site contains proxies for scales that are sometimes used to detect biased responding (e.g., Paulhus (1991) Self-Deception and Impression-Management scales), the site does not provide a unique set of new general-purpose IPIP validity indices.

2.6. Warnings for nonprofessionals

The first link on the IPIP front page leads to a warning to job applicants seeking information that they believe will help them "cheat" on personality tests used in personnel selection and placement. They are first informed that the purpose of selection instruments is to find the best match between individuals and jobs, so that trying to cheat is therefore self-defeating. They are next told that most personality inventories include items specifically

devised to detect "cheaters." Finally, they are informed that there is no evidence whatsoever that using the IPIP website will in any way enhance the likelihood that one will be selected for a job.

Hundreds of persons surf the Web each day in search of psychological tests to complete for their amusement or self-insight. At the IPIP site, an Overview page informs visitors that the site is for research purposes only, and that no personality measures are administered there. Persons interested in completing an inventory to receive feedback are directed to a site whose express purpose is to provide this educational service: http://www.personal.psu.edu/~j5j/IPIP/.

The IPIP Web site does not contain detailed instructions on how to assemble items into scales, administer and score scales, or interpret results, although suggested instructions and sample questionnaires are provided. The Site Overview page warns individuals who have not completed a university course or two in psychological assessment that the site contains highly technical scientific information. John A. Johnson is listed as a contact for limited consulting help, and some of his experiences in this capacity are reported below.

3. A consultant's experiences with the IPIP: The good, the bad, and the ugly

From the time he began serving as the IPIP consultant in August 1999 through January, 2005, Johnson (2005b) had received approximately 800 e-mail messages from over 400 individuals interested in using IPIP scales for research or education. These individuals included high school students with science-fair projects, high school teachers creating demonstrations for their courses, graduate students conducting dissertation research, college and university faculty desiring to employ IPIP scales in their teaching or research, computer programmers interested in setting up an assessment Web site, and human-resource professionals looking for scales to use in personnel selection. In addition, Johnson has received over 400 messages from approximately 250 persons who had completed one of his online IPIP inventories and had questions or comments about that experience. These figures, which do not include individuals who simply downloaded IPIP items or scales without contacting the consultant, have been increasing steadily and indicate considerable interest in the IPIP as a resource. Comments from IPIP users reveal a number of features that account for the IPIP's growing popularity. An analysis of these features suggests that they hold potential for both positive and negative consequences for personality measurement.

First of all, the IPIP is free. The cost of commercial inventory booklets and answer sheets may be trivial to researchers holding large grants at major research universities, but not so for researchers—particularly students—at smaller colleges, and especially when they wish to use large samples (Ashton, 2005; Cloninger, 2005). Furthermore, scoring a commercial inventory may involve an additional charge. Thus, the overall cost of commercial inventories can be prohibitive for many researchers. By providing free resources, the IPIP helps to level the playing field between researchers with means and those without.

This is not to say that all commercial companies are indifferent to less affluent researchers. Some companies offer special discount rates to students, and some may donate testing materials and scoring services to researchers (Gough, 2005). Indeed, some may support both student and faculty research at no or low cost. At least one test company provides stipends for student support and prizes for student competition, conducts data analyses and writes reports for individuals, all to promote the cause of good personality research

world-wide (Hogan, 2005). This assistance could be viewed quite positively in cases where the inventory users lack sufficient expertise and therefore require intellectual guidance as well as material support. On the other hand, the assistance that a company provides might be seen by some as unduly restrictive, as when a company scores answer sheets and returns reports instead of releasing scoring keys, which can prevent some kinds of research.

A second attractive feature of the IPIP is the speed with which its items and scales can be obtained. Whereas commercial publishers require a time-consuming, formal application to establish user qualifications and then consume additional time by shipping inventories, answer sheets, or computer disks, the IPIP site provides scales via the Internet with no restrictions and no delay. Although this is clearly a boon from a user's viewpoint, the immediate and unrestricted availability of IPIP materials does raise questions about possible misuse by nonprofessionals. Johnson (2005b) discussed the following message as an example of a user who, however well intentioned, may not know how to score IPIP items correctly and may misinterpret IPIP scale scores:

> "I'm a teacher in a high school and I wanted to administer the Big Five questionnaire that is on the website. However, I'm confused about how to score it. I see that every question is associated with each of the 5 factors, and it has a plus or minus associated with it as well. However, I am then a little confused about how to take the five potential answers a person could give to each question, incorporate the +/− and arrive at a score/percentile that I could use."

Johnson (2005b) noted that this individual, by sending an inquiry to the consultant, had at least a chance of learning how to score and interpret the scales correctly. Indeed, consulting support can be crucial to proper use of the IPIP (Cloninger, 2005). Who knows how many persons who did not bother to contact the consultant might be misusing IPIP measures?

The clear visibility of over 2000 personality-related items constitutes a third attractive feature of the IPIP. Whereas new investigators who do not possess a copy of a commercial inventory must rely on second-hand descriptions of its item content, any investigator can examine the content and the wording of every item on the IPIP Web site to ensure that the items are appropriate for the population being assessed. The transparency of item content enables IPIP users to distinguish among nuanced versions of a construct—for example, whether a "positive affect" scale reflects hearty, energetic emotions such as *cheerfulness* and *enthusiasm* or mild, gentle feelings such as *warmth* and *sympathy* (Johnson & Ostendorf, 1993).

Another advantage related to the visibility of the items, coupled with the wide range of constructs measured by one or more IPIP scales, is that investigators can target particular constructs of interest to them, using scales specifically relevant to their research goals. Thus, they are not constrained by the limited number of scales on any single commercial inventory (Ashton, 2005; Johnson, 2005b).

The visibility of the scoring keys for all IPIP scales is a fourth attractive feature because it allows researchers to conduct item-level analyses. Knowing the specific items that are included in each scale is necessary for computing coefficient-α reliability estimates and for verifying whether items are loading properly in an item-level factor analysis. Although some companies might be willing to compute reliability statistics for a researcher's sample, turning over all of the analyses of one's data to any company disallows the interactive, iterative procedures optimal for scale development. With the IPIP, one can begin with

standard keying and then experimentally remove and add items to examine the effect on internal consistency in one's sample. As such, the IPIP provides training opportunities for students learning scale-development techniques (Ashton, 2005) as well as possibilities for professional scale development (Buchanan et al., 2005).

A fifth attractive feature of the IPIP is that the non-copyrighted items afford a degree of flexibility not found in commercial inventories. Users can mix and match items and scales according to their needs. They can present the items in any order they choose. They can reword items and translate them into other languages without asking permission. They can create Web-based versions of IPIP scales and reap all of the benefits of that mode of data collection (Gosling et al., 2004). The downside of the freedom to tinker with public-domain scales is that results from studies using different versions of a scale become incommensurable. Johnson (2005b) used, as a non-IPIP example, the following excerpt from a message referring to Goldberg (1992) public-domain markers for the Big-Five factor structure:

> "To review, we are working with a large dataset […]. The official information claims that measures of Neuroticism and Extroversion were derived from your 1992 scales […] Unfortunately, when we looked at the specific items, they were only loosely related to the scales in your 1992 article."

Cumulative scientific progress requires enough methodological uniformity to compare findings from different studies. The potential for IPIP scales to mutate in unforeseeable ways may undermine their requisite uniformity for scientific progress (Hogan, 2005; Johnson, 2005b).

3.1. A potential danger: The uncertain relations between IPIP scales and their parent scales

As already noted, there is a rapidly expanding literature reporting scientific studies that have included IPIP scales. Interestingly, many of these studies have used IPIP measures of constructs that are already in the public domain, such as the 50-item IPIP inventory targeted at Goldberg (1992) set of 100-adjective markers of the lexical Big-Five factor structure. Another popular IPIP measure is the 50-item IPIP inventory targeted at the five domain constructs included in the commercial NEO-PI-R. In the former case, investigators are using one public-domain measure in place of another, perhaps because of a preference for IPIP items over single trait-descriptive adjectives. In the latter case, investigators are opting for a public-domain inventory over a commercial one. In both cases, one must worry about the extent to which IPIP measures are equivalent to their parent scales.

The IPIP resembles open-source software, at least in concept, because many of its scales were designed to provide public-domain alternatives to commercial products (Hogan, 2005). To some extent, therefore, interest in the IPIP could be tied to its ability to mimic commercial scales. That is, if commercial scales had never been developed (at a cost that had to be compensated by sales—Eber, 2005), the IPIP would be less useful. Even though the IPIP is an item pool that can be used to generate entirely novel scales, the visibility and success of the IPIP may be dependent on its ability to measure, with high fidelity, the constructs represented in commercial inventories.

Because the items selected for IPIP proxies of commercial scales are based on empirical correlations with the original scales, the correlations between the proxy and parent scales tend to be high. For example, the mean correlation between the 30 facet scales of the NEO-PI-R (Costa & McCrae, 1992) and the corresponding IPIP scales is .73 (.94 after

correcting for attenuation due to unreliability—Goldberg, 1999a). The substantial overlap between IPIP and parent scale variance might lead one to conclude that the corresponding scales are measuring the same constructs, but that conclusion could be incorrect (Ashton, 2005). For example, in a large internet sample three of 30 facet scale scores from the IPIP representation of the NEO-PI-R did not show their highest loading on the expected domain factor, and two items from the IPIP 50-item measure of the Big Five did not show their highest loading on the expected factor (Buchanan et al., 2005). And, a content analysis indicated some differences in the content of the NEO-PI-R Openness to Values facet scale and the IPIP scale meant to represent that construct (Johnson, 2001).

As another example of possible construct differences, Gough (2005) contrasted the Dominance (Do) scale from the CPI (Gough & Bradley, 1996) with the corresponding IPIP proxy. Gough noted that there are at least two variants of the disposition toward dominance, one stemming from self-aggrandizing motives, and the other from pro-social drives toward constructive and consensual goals. The CPI Do scale was designed to minimize any connection with the self-aggrandizing theme, and to maximize a connection with pro-social motives and behavior. Two examples of Do scale items (both scored "true") illustrate this point: "Every citizen should take the time to find out about national affairs, even if it means giving up some personal pleasures," and "When the community makes a decision, it is up to a person to help carry it out even if he or she had been against it."

On the other hand, the IPIP proxy for Do contains items such as: "Impose my will on others," "Try to outdo others," and "Demand explanations from others." The IPIP scale, unlike the CPI Do scale, appears to be more aligned with the self-aggrandizing variant of dominance, perhaps because of a lack of pro-social dominance items in the IPIP item pool at the time that this scale was developed. Thus, even if the IPIP and CPI dominance scales correlate highly, the psychological constructs they represent may differ in significant ways. Whether these differences are important must be judged ultimately on the external correlates of item responses rather than casual inspection of item content (Gough, 2005; Hogan & Nicholson, 1988; Johnson, 1997).

The construct validities of scales on some of the major commercial inventories are supported by decades of research. It would be convenient if the overlap between these scales and the IPIP versions of these scales allowed one to assume that the IPIP scales will show the same correlations with external criteria. That assumption is not always warranted, and equivalences and differences between constructs tapped by commercial scales and their IPIP proxies can be discovered only by conducting comparative-validity studies. What is important to realize is that the IPIP measure of a construct included in some personality inventory could turn out to be more valid in diverse contexts than the original scale itself. Only further research will tell.

3.2. The partially fulfilled promise of the IPIP as a collaboratory

It is too early to conclude whether the IPIP is actually accelerating scientific progress in personality research or whether the concerns expressed by some critics (that unqualified users will cause harm by misusing the IPIP and/or that unrestricted revisions of items and scales will cause fragmentation rather than unification) will detract from the personality enterprise. Nonetheless, the growing popularity of the IPIP as a source of items and scales suggests that this project is off to a promising start. Yet Goldberg's (1999a) vision for the IPIP project went beyond providing a pool of items and scales. The IPIP Web site was

envisioned as a nexus that connects researchers from around the world for collaborative personality research. Researchers could provide data sets containing scores from IPIP measures for reanalysis by other researchers, and the Web site could serve as an interactive forum for scientific discussions. To what degree is this sort of collaboration and data sharing occurring?

The Oregon Research Institute (ORI) currently maintains a data set from a sample of approximately 750 home owners in the Eugene–Springfield (Oregon) community who have completed all of the IPIP items and a large set of other psychological measures (Goldberg, 2005b). Because this data set contains scores from many of the major commercial personality inventories, as well as IPIP measures, it has been lauded for its ability to demonstrate overlap and uniqueness across different proprietary instruments (Ashton, 2005; Cloninger, 2005) and facilitate communication between researchers committed to different measurement traditions (Eber, 2005). Although the Eugene–Springfield data set cannot be downloaded directly from the IPIP Web site, ORI routinely provides data from this archive to researchers upon request. ORI also considers requests from researchers who would like a particular measure administered to the Eugene–Springfield community sample. This represents a step toward the IPIP-as-collaboratory vision.

For the ultimate in collaborative interactivity, Johnson (2005b) proposed a dynamic data set in which an individual researcher or research team initiates data collection by having research participants complete IPIP measures on-line. As data collection progresses, other researchers could post additional on-line measures to be completed by the participants who responded to the original measures. The result would be similar to ongoing longitudinal studies such as those conducted by Costa and McCrae (1993), except that both the subject pool and the investigative team would include participants from around the world. To fulfill its potential as a collaboratory, the IPIP Web site should provide immediate access not only to ORI data but also to data sets from other researchers.

Will this idea fly? To see, stay tuned to the IPIP site over the next 10 years.

References

Ashton, M. C. (2005). How the IPIP benefits personality research and education. In L. R. Goldberg (Chair), *The international personality item pool and the future of public-domain personality measures. Presidential symposium at the sixth annual meeting of the association for research in personality,* New Orleans, January.

Ashton, S. G., & Goldberg, L. R. (1973). In response to Jackson's challenge: The comparative validity of personality scales constructed by the external (empirical) strategy and scales developed intuitively by experts, novices, and laymen. *Journal of Research in Personality, 7*, 1–20.

Barchard, K. A. (2001). Emotional and social intelligence: Examining its place in the nomological network. Unpublished Doctoral Dissertation, Department of Psychology, University of British Columbia, Vancouver, BC, Canada.

Buchanan, T., Johnson, J. A., & Goldberg, L. R. (2005). Implementing a five-factor personality inventory for use on the Internet. *European Journal of Psychological Assessment, 21*, 116–128.

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*, 319–333.

Cloninger, C. R. (1994). *The Temperament and Character Inventory (TCI): A guide to its development and use*. Washington University, St. Louis, MI: Center for the Psychobiology of Personality.

Cloninger, C. R. (2004). *Feeling good: The science of well-being*. New York: Oxford University Press.

Cloninger, C. R. (2005). Comments on Lew Goldberg's IPIP. In L. R. Goldberg (Chair), *The international personality item pool and the future of public-domain personality measures. Presidential symposium at the sixth annual meeting of the association for research in personality,* New Orleans, January.

Conn, S. R., & Rieke, M. L. (1994). *The 16PF fifth edition technical manual.* Champaign, IL: Institute for Personality and Ability Testing.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R™) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

Costa, P. T, Jr., & McCrae, R. R. (1993). Psychological research in the Baltimore longitudinal study of aging. *Zeitschrift für Gerontologie, 26*, 138–141.

Costa, P. T., Jr., & McCrae, R. R. (1997). Stability and change in personality assessment: The revised NEO Personality Inventory in the year 2000. *Journal of Personality Assessment, 68*, 86–94.

Costa, P. T., Jr., & McCrae, R. R. (1999). Reply to Goldberg. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 29–31). Tilburg, The Netherlands: Tilburg University Press.

Eber, H. (2005). IPIP: A milestone on the path to conceptual clarity. In L. R. Goldberg (Chair), *The international personality item pool and the future of public-domain personality measures. Presidential symposium at the sixth annual meeting of the association for research in personality,* New Orleans.

Finholt, T. A., & Olson, G. M. (1997). From laboratories to collaboratories: A new organizational form for scientific collaboration. *Psychological Science, 8*, 28–36.

Goldberg, L. R. (1972). Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monograph, 7*, No. 72-2.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*, 26–42.

Goldberg, L. R. (1995). What the hell took so long? Donald Fiske and the Big-Five factor structure. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A Festschrift honoring Donald W Fiske* (pp. 29–43). Hillsdale, NJ: Erlbaum.

Goldberg, L. R. (1999a). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 7–28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R. (1999b). Response to Costa and McCrae. In I. Mervielde, I. Deary, F. Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe* (Vol. 7, pp. 33–34). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. R. (2005a). A scientific collaboratory for the development of advanced measures of personality traits and other individual differences. In L. R. Goldberg (Chair), *The international personality item pool and the future of public-domain personality measures. Presidential symposium at the sixth annual meeting of the association for research in personality,* New Orleans, January.

Goldberg, L. R. (2005b). The Eugene–Springfield community sample: Information available from the research participants. *ORI Technical Report, 45*(1).

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*, 93–104.

Gough, H. G. (2005). Remarks for the ARP presidential symposium. In L. R. Goldberg (Chair), *The international personality item pool and the future of public-domain personality measures. Presidential symposium at the sixth annual meeting of the association for research in personality,* New Orleans, January.

Gough, H. G., & Bradley, P. (1996). *CPI manual* (third ed.). Palo Alto, CA: Consulting Psychologists Press.

Hendriks, A. A. J. (1997). *The construction of the Five-Factor Personality Inventory (FFPI)*. Groningen, The Netherlands: Rijksuniversiteit Groningen.

Hofstee, W. K. B., de Raad, B., & Goldberg, L. R. (1992). Integration of the Big-Five and circumplex approaches to trait structure. *Journal of Personality and Social Psychology, 63*, 146–163.

Hogan, R. (2005). On the IPIP. In L. R. Goldberg (Chair), *The international personality item pool and the future of public-domain personality measures. Presidential symposium at the sixth annual meeting of the Association for Research in Personality,* New Orleans, January.

Hogan, R., & Hogan, J. (1992). *Hogan personality inventory manual.* Tulsa, OK: Hogan Assessment Systems.

Hogan, R., & Nicholson, R. A. (1988). The meaning of personality test scores. *American Psychologist, 43*, 621–626.

Jackson, D. N. (1994). *Jackson personality inventory-revised manual.* Port Huron, MI: Sigma Assessment Systems.

Jackson, D. N., Paunonen, S. V., & Tremblay, P. F. (2000). *Six factor personality questionnaire manual.* Port Huron, MI: Sigma Assessment Systems.

Johnson, J. A. (1997). Seven social performance scales for the California Psychological Inventory. *Human Performance, 10*, 1–30.

Johnson, J. A. (1999a). Persons in situations: Distinguishing new wine from old wine in new bottles. In I. Van Mechelen & B. De Raad (Eds.), *Personality and situations [Special Issue]. European Journal of Personality, 13*, 443–453.

Johnson, J. A. (1999b). *Some hypotheses concerning attempts to separate situations from personality dispositions.* Invited paper presented at the Sixth European Congress of Psychology, Rome.

Johnson, J. A. (2000). Predicting observers' ratings of the Big Five from the CPI, HPI, and NEO-PI-R: A comparative validity study. *European Journal of Personality, 14*, 1–19.

Johnson, J. A. (2001). *Screening massively large data sets for non-responsiveness in web-based personality inventories.* Invited talk to the joint Bielefeld-Groningen Personality Research Group, University of Groningen, The Netherlands, May. Available at: <http://www.personal.psu.edu/~j5j/papers/screening.html>.

Johnson, J. A. (2005a). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality, 39*, 103–129.

Johnson, J. A. (2005b). The good, the bad, and the ugly: The IPIP consultant's experiences recounted. In L. R. Goldberg (Chair), *The international personality item pool and the future of public-domain personality measures. Presidential symposium at the sixth annual meeting of the association for research in personality,* New Orleans, January.

Johnson, J. A., & Ostendorf, F. (1993). Clarification of the five-factor-model with the abridged big five dimensional circumplex. *Journal of Personality and Social Psychology, 65*, 563–576.

Kurtz, J. E., & Parrish, C. L. (2001). Semantic response consistency and protocol validity in structured personality assessment: The case of the NEO-PI-R. *Journal of Personality Assessment, 76*, 315–332.

Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research, 39*, 329–358.

Ones, D. S., & Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance, 11*, 245–271.

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). New York: Academic Press.

Piedmont, R. L., McCrae, R. R., Riemann, R., & Angleitner, A. (2000). On the invalidity of validity scales: Evidence from self-reports and observer ratings in volunteer samples. *Journal of Personality and Social Psychology, 78*, 582–593.

Saucier, G. (1997). Effects of variable selection on the factor structure of person descriptors. *Journal of Personality and Social Psychology, 73*, 1296–1312.

Tellegen, A. (in press). *MPQ (Multidimensional Personality Questionnaire): Manual for administration, scoring, and interpretation.* Minneapolis, MN: University of Minnesota Press.