



The Case for Noncognitive Assessments

Patrick C. Kyllonen

ETS is known for its work on the SAT[®], Graduate Record Examinations[®] (GRE[®]), National Assessment of Educational Progress, The Praxis Series[™], and other tests of knowledge and cognitive ability.

But does ETS have assessments of *noncognitive* qualities¹—persistence, dependability, motivation, the ability to work with others, intercultural sensitivity? Do these matter? Do they affect success in school or in the workplace?

Are Noncognitive Skills Important?

They apparently are important in industry. Employers report valuing job stability and dependability, and they often use noncognitive assessments in employee hiring decisions, for good reason. Meta-analyses (analyses of the combined results from multiple studies) have shown that noncognitive measures provide a 20% improvement over cognitive ability measures in predicting training success and job performance (Schmidt & Hunter, 1998).

In education the picture is a bit murkier because noncognitive measures are rarely used to assess students. Still, research has shown that noncognitive factors predict grades in K-12, as well as social outcomes (Caprara, Barbanelli, Pastorelli, Bandura, & Zimbardo, 2000). And in higher education, faculty members believe that noncognitive variables are important determinants of school success. Faculty value noncognitive qualities, such as persistence, tenacity, collegiality,

communication, and enthusiasm, as much as cognitive qualities, such as research experience and mastery of discipline, for admissions and as desirable outcomes (Kyllonen, Walters, & Kaufman, in press; Walpole, Burton, Kanyi, & Jackenthal, 2002).

There is other evidence for the importance of noncognitive skills. The Nobel Laureate Economist James Heckman found that General Educational Development (GED) recipients (half of all high school dropouts) score as highly on cognitive tests as noncollege-bound high school graduates and higher than other dropouts (Heckman & Rubenstein, 2001). They earn more than other dropouts, are less likely to leave military service, and are more likely to go back to school. However, their advantage over other dropouts seems due to their cognitive ability advantage, because when statistically controlling for cognitive ability, GED recipients actually earn significantly less, are more likely to leave military service, and are less likely to go back to school than non-GED dropouts. Heckman and Rubenstein suggest that GED certification thus is a “mixed signal” indicating higher cognitive, but lower noncognitive skills. As they put it, “Inadvertently, a test has been created that separates out bright but nonpersistent and undisciplined dropouts from other dropouts.... GEDs ... lack the abilities to think ahead, to persist in tasks, or to adapt to their environments” (p. 146).

The importance of noncognitive skills also can be seen in studies of early childhood interventions such as Head Start and the Abecedarian project. Historically, evaluations have focused on cognitive-test-score outcomes (e.g., Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002). But it may be that the major outcomes of these kinds of programs are noncognitive—the development of the qualities that lead to less criminal behavior and higher employment, earnings, and high school graduation rates (Schweinhart et

¹ The term *noncognitive*, although a misnomer, is widely used in psychology and measurement. Other relevant terms are *nonacademic*, *socioaffective*, *affective-motivational*, and *personality*.

al, in press). The Big Brothers Big Sisters program, a noncognitive intervention, has been shown to increase school success, reduce drug and alcohol involvement, and lead to better relationships with parents (Tierney, Grossman, & Resch, 1995).

The Gap

An argument for noncognitive assessments is that they go beyond “academic intelligence” as Robert Sternberg (1985) puts it and tap the full range of qualities that affect and are affected by schooling. But another argument is that using noncognitive assessments may serve to reduce the test score gap, the mean difference in scores between White and Black test takers commonly observed on more narrowly focused cognitive assessments. Research suggests that there is no score gap or a reduced score gap on noncognitive assessments (Sackett, Schmitt, Ellingson, & Kabin, 2001). Combining noncognitive and cognitive test scores in a selection index would result in a reduced overall score gap.

It may also be that noncognitive factors, such as study skills and attitudes toward learning, contribute to the score gap. Research on this issue has been inconsistent (Jencks & Phillips, 1998), but if there is merit to this idea, noncognitive interventions may address gap issues.

Taken together, these findings suggest that noncognitive factors are important and ought to be assessed. What is ETS’s track record with noncognitive assessments?

History

From its beginnings, ETS recognized the importance of noncognitive skills and sought ways to promote noncognitive assessments. In 1948, ETS’s first president, Henry Chauncey, discussed the “Census of Abilities,” a logical next step beyond the SAT, to assess

... personal qualities, some of which may be drive (energy), motivation (focus of energy), conscientiousness, intellectual stamina . . . ability to get along with others.... interests, such as aesthetic, religious, abstract, social, economic, political, manipulative (Lemann, 1995, p. 84)

This led to ETS establishing the personality research group, at one point (1959-1967) headed

by Samuel Messick. The group and contributors, including Lawrence Stricker, Nathan Kogan, Irving Sigel, and Douglas Jackson, conducted wide-ranging research on noncognitive topics throughout the 1960s—attitudes, creativity, and acquiescence and social desirability as response styles on personality scales (e.g., Messick, 1996). Also during that time, ETS established a contract with Isabel Briggs Myers for the Myers-Briggs Type Indicator, now the largest personality inventory in use with two million copies sold every year. ETS did not pursue that work because of concerns about the validity and practical application of the test (Stricker & Ross, 1964).

Other significant ETS forays into noncognitive research have included cognitive styles (Messick, 1996), field-dependence (Witkin & Goodenough, 1981), in-basket testing, now widely used in industry (Frederiksen, Saunders, & Wand, 1957), documented accomplishments (Baird, 1979; Stricker, Rock, & Bennett, 2001), and more recently, stereotype threat (Stricker & Bejar, 2004; Stricker & Ward, 2004; Walters, Lee, & Trapani, 2004). However, outside of a small handful of assessments—such as the widely used SIGI PLUS student guidance system² based on interests (Katz, 1993)—noncognitive research did not result in operational assessments.

Measurement Issues

With a strong justification for developing noncognitive assessments and ETS’s history of involvement with them, why does the organization not offer a full array of noncognitive assessments today? Why is there no noncognitive GRE or SAT subtest? The answer: Many policy makers and scientists are skeptical that noncognitive qualities can be measured reliably and in a valid way. Typically, in both research and operational use, noncognitive qualities are assessed through self-ratings. Examinees are asked questions such as, “Are you exacting in your work?” “Do you get chores done right away?” “Do you keep your emotions under control?” “Do you take time to reflect on things?” There are two problems with these kinds of ratings—the standard is not clear (i.e., relative to whom?), and they are easily

² ETS developed SIGI PLUS, which is now available through Valpar International.

faked. In almost any serious discussion of the use of noncognitive assessments, the issue of “fakability” or “coachability” comes up, and this issue is the trump card that thwarts further discussion.

But does it have to?

Considerable thought has been devoted to ways to get around the fakability problem in order to be able to use noncognitive assessments.

Fake-Resistant Noncognitive Assessments

At least since the 1960s, psychologists have recognized the seriousness of the fakability threat and have investigated assessments that are more like ability tests (Cattell & Scheier, 1960). Areas currently being actively researched under this heading include the following.

Reaction Time Measures. The more familiar something is to us, the quicker we recognize it, and that’s the principle behind the use of reaction time to measure personality. The simplistic idea that we should be quicker to recognize trait terms that describe us has not panned out. But a variant on this idea—the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998)—may yield more fruitful results. In the IAT, reaction time reflects the naturalness of an association for a person between two objects, for example, oneself and a trait term. The “emotional stroop” test (Williams & Nulty, 1986) also fits into this category.

Emotional Intelligence Measures. Current performance tests of emotional intelligence, such as the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT), present a suite of objective noncognitive assessments. Examples are Blends (“What emotion results from combining fear and anger?”), and Faces (“How surprised is the person pictured? How happy?”). Paul Ekman (2003) has developed a test of one’s ability to determine the authenticity of actors’ expressions of various emotions, and Klaus Scherer and colleagues (Scherer, Banse, & Wallbott, 2003) have done something similar with vocal utterances. ETS is currently investigating these and other such measures (Schulze & Roberts, 2005).

Situational Judgment Tests. These tests describe a scenario posing some kind of problem, and the examinee is asked how best to solve the problem. The format permits both text and video-based

tests, it has been researched widely (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001) and is now commonly used in industry. It has the advantage of enabling measurement of a wide variety of constructs, and situations can mimic real life circumstances, providing face validity. Research has shown that these tests can measure noncognitive as well as cognitive qualities (McDaniel & Nguyen, 2001).

Biodata. Resumes are almost universally used in the employment and educational application process, and there has been considerable research on the reliability and validity of various ways of capturing resume data. A finding is that data that is verifiable at least in principle may be more valid and less subject to faking.

Forced Choice. An examinee is presented two or more noncognitive descriptors (e.g., punctual, agreeable) and must indicate which descriptor describes him or her better. This format avoids the obvious susceptibility to faking inherent in conventional self-ratings, particularly if descriptors are of equal social desirability.

Multiple Measures. The only behavioral indicators typically examined for any kind of psychological or educational assessments are accuracy and response time. But other measures such as confidence, stress (e.g., heart rate), and eye movements have been researched. Some of these alternative indicators may serve to provide a richer description of the examinee’s noncognitive as well as cognitive state.

Others’ Ratings

If the problem of self-ratings is fakability, then why not use ratings by others, such as teachers or advisors? After all, isn’t this what a letter of recommendation is? This idea motivated ETS’s development of a Standardized Letter of Recommendation (SLR), which is being used for selecting ETS summer interns and fellowship recipients. It also is in pilot testing in various university settings. The advantage of the SLR is that it specifies the full range of valid dimensions, requests quantitative as well as qualitative ratings, and does it all with Internet convenience (Kyllonen & Kim, 2005; Walters et al, 2004). The system also allows systematic retention (by the institution) of historical records for studies and analyses.

Low-Stakes Assessments

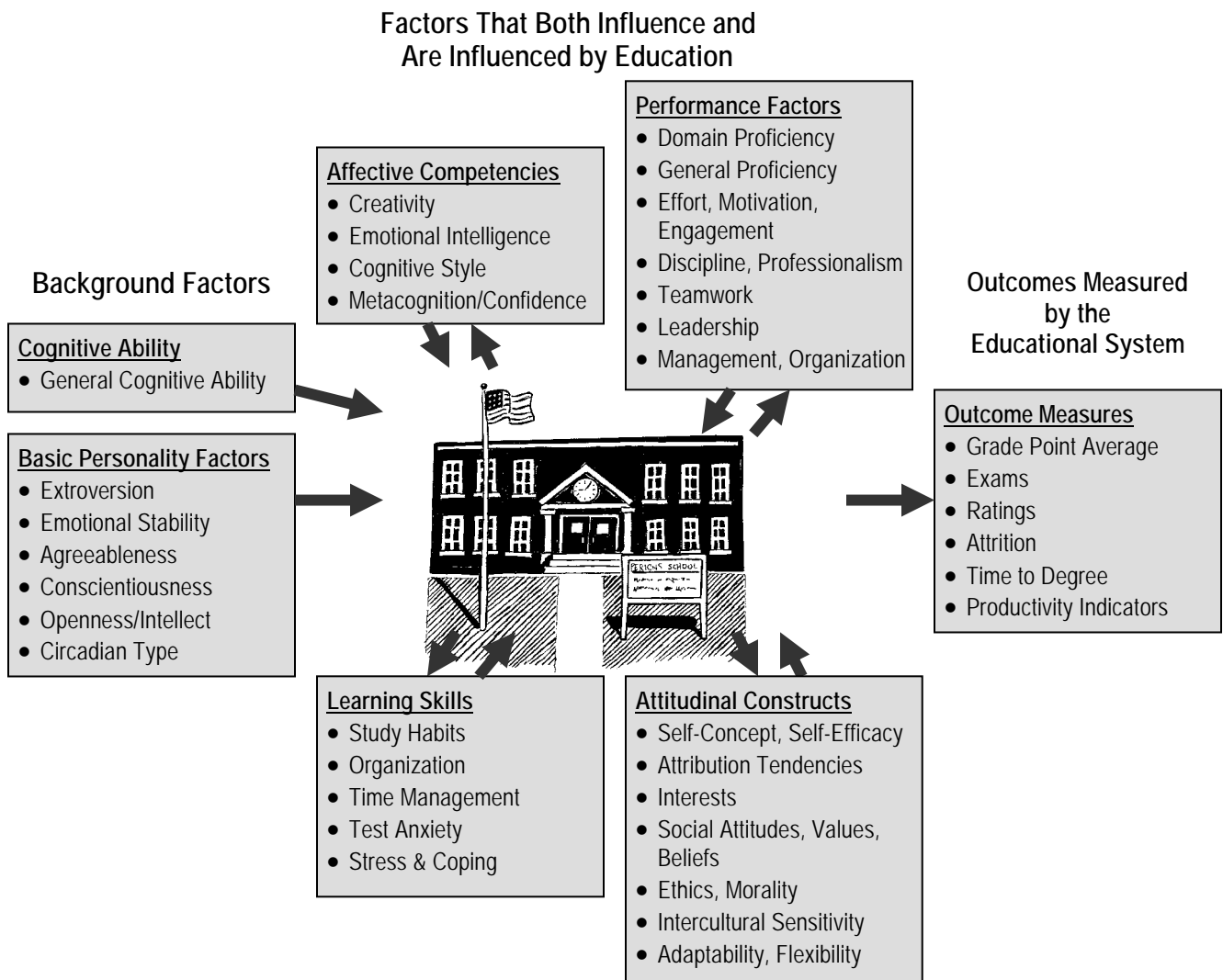
Noncognitive assessments do not have to be used in a high-stakes setting to have an effect on education. Career guidance systems such as SIGI PLUS use noncognitive assessments to help students understand themselves better so they make better educational decisions, such as selecting an undergraduate major. One can imagine a smorgasbord of self-help assessments covering topics ranging from test anxiety to time management, circadian rhythms, values, beliefs, and attitudes. A student could profitably complete such a battery as part of a self-diagnosis effort. For example, research suggests that some attitudes towards tests—such as the attitude that low scores are due to low effort rather than low ability—are more productive than others. A self-assessment

could reveal these kinds of counterproductive attitudes and suggest readings and exercises to overcome them.

What Are the Noncognitive Constructs?

Figure 1 depicts a context for whole person assessment in education, which also identifies important noncognitive factors that ETS is currently investigating. The key point is that a wide range of noncognitive factors—attitudes, learning skills, performance factors, and affective competencies—both affect students’ educational experience and outcomes and are affected by education. Noncognitive assessments therefore can play multiple roles—admissions, placement, diagnosis, outcomes, institutional studies, and others.

Figure 1. A Context for Whole Person Assessment in Education



Current Work and Next Steps

Higher Education

In addition to the SLR (described above), ETS has recently developed several noncognitive assessments and is experimenting with many more. The organization's Center for New Constructs has developed situational judgment tests for selecting business analysts (Kyllonen & Lee, 2005), as well as a video-based noncognitive assessment for measuring communication skills of medical college applicants (Kyllonen, 2005). The center has developed prototypes of situational judgment tests for college, graduate school, and professional school admissions and is currently exploring opportunities to validate these. In addition, the center is pilot-testing 28 noncognitive assessments on various college campuses with results expected within the next year.

Community Colleges

The needs of community colleges can be quite different from those of other higher education establishments. From discussions with community college leaders, ETS has found that there is a major need for a self-help Web site that provides useful information to the student, particularly the at-risk student, about issues important for achieving that student's educational goals. The Center for New Constructs has developed such a Web site. It provides help in exploring career goals and monitoring progress and also provides self-assessments on topics ranging from test anxiety to circadian rhythms, confidence, attitudes, and values, with links to further readings and other interventions to help students overcome perceived problems.

K-12

Because of the No Child Left Behind legislation, schools today are most concerned with meeting their adequate yearly progress (AYP) targets in math and reading. But research (Organisation for Economic Cooperation and Development, 2004) has clearly shown links between noncognitive variables—such as student engagement, learning skills, and school climate—and academic achievement outcomes. ETS currently is pilot-testing a system of assessing these noncognitive variables in several urban and suburban K-12 districts and developing intervention and

monitoring plans to help districts achieve their AYP targets by improving their noncognitive skills and qualities.

The center also is attempting to explore the use of noncognitive assessments in several teacher quality initiatives.

Other Applications

ETS is currently developing and field-testing a wide variety of noncognitive assessments for a variety of uses besides the ones mentioned above. These include selection, diagnosis, and outcomes assessment for industry, the military, and international higher education institutions.

Summary

A graduate school dean once stated that his program measured noncognitive qualities with three objective measures: how far the applicant would be traveling to attend school (cross-country is very good), whether the applicant would be leaving a job (giving up a job is a sign of seriousness of purpose), and whether the applicant would be bringing a family (uprooting a family indicates something other than a casual choice). According to the dean, an affirmative to these three questions portends persistence through degree completion. These items are appealing because they are perceived to measure an important noncognitive quality, that quality is believed to relate to an important outcome variable (degree completion), and the measures themselves cannot be faked, at least not easily. There are obvious fairness problems with these measures (e.g., what about the dedicated applicant who happens to be single and live around the corner?), but their appeal is understandable. ETS's hope for its noncognitive initiative is to develop assessments that provide the benefits but without the drawbacks of the dean's three measures. ETS continues actively to pursue the development of noncognitive assessments, to explore a variety of possible uses for different educational segments ranging from K-12 to graduate school, and to demonstrate how use of such assessments can affect student outcomes and educational conversations in a productive way.

References

- Baird, L. L. (1979). *Development of an inventory of documented accomplishments for graduate admissions* (GRE Board Research Rep. No. 77-3R). Princeton, NJ: ETS.
- Campbell, F. A., Ramey, C. T., Pungello, E. P., Sparling, J., & Miller-Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian project. *Applied Developmental Science, 6*, 42-57.
- Caprara, G.V., Barbanelli, C., Pastorelli, C., Bandura, A., & Zimbardo, P.G. (2000). Prosocial foundations of children's academic achievement. *Psychological Science, 11*, 302-306.
- Cattell, R. B., & Scheier, I. H. (1960). *Handbook and test kit for the Objective-Analytic (O-A) Anxiety Battery*. Champaign, IL: Institute for Personality & Ability Testing.
- Ekman, P. (2003). *Emotions revealed: Recognizing faces and feelings to improve communication*. New York: Times Books, Henry Holt and Company.
- Frederiksen, N., Saunders, D., & Wand, B. (1957). The in-basket test. *Psychological Monographs, 71*(9, Whole No. 438).
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*, 1464-1480.
- Heckman, J. J., & Rubenstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review, 91*(2), 145-149.
- Jencks, C., & Phillips, M. (Eds.). (1998). *The black-white test score gap*. Washington, D.C.: Brookings Institution Press.
- Katz, M. R. (1993). *Computer-assisted career decision making: The Guide in the machine*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kyllonen, P. C. (2005). Video-based communication skills test for use in medical college. In N. Gafni (Organizer), *Assessment of non-cognitive factors in student selection for medical schools*. Symposium conducted at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Kyllonen, P. C., & Kim, S. (2005). Personal qualities in higher education: Dimensionality of Faculty ratings of students applying to graduate school. In *Dimensionality and model fit*. Session presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Kyllonen, P. C., & Lee, S. (2005). Assessing problem solving in context. In O. Wilhelm & R. Engle (Eds.), *Handbook of understanding and measuring intelligence*. Thousand Oaks, CA: Sage Publications, Inc.
- Kyllonen, P. C., Walters, A., & Kaufman, J. (in press). *Noncognitive constructs in graduate education* (GRE Board Report No. 00-11R). Princeton, NJ: ETS.
- Lemann, N. (1995, September). The great sorting. *Atlantic Monthly, 276*(3), 84-100.
- McDaniel, M. A., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of situational judgment tests to predict job performance: A clarification of the literature. *Journal of Applied Psychology, 86*, 730-740.
- McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment, 9*, 103-113.
- Messick, S. (1996). Human abilities and modes of attention: The issue of stylistic consistencies in cognition. In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 77-96). Hillsdale, NJ: Erlbaum.
- Organisation for Economic Cooperation and Development. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: Author.
- Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative action world. *American Psychologist, 56*, 302-318.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32*(1), 76-92.
- 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.
- Schulze, R., & Roberts, R. D. (Eds.). (2005). *Emotional intelligence: An international handbook*. Cambridge, MA: Hogrefe & Huber Publishers.
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (in press). *Lifetime effects: The High/Scope Perry Preschool study through age 40* (Monographs of the High/Scope Educational Research Foundation, 14). Ypsilanti, MI: High/Scope Press.
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.
- Stricker, L. J., & Bejar, I. I. (2004). Test difficulty and stereotype threat on the GRE general test. *Journal of Applied Social Psychology, 34*(3), 563-597.
- Stricker, L. J., Rock, D. A., & Bennett, R. E. (2001). Sex and ethnic-group differences on accomplishments measures. *Applied Measurement in Education, 14*(3), 205-218.

-
- Stricker, L. J., & Ross, J. (1964). An assessment of some structural properties of the Jungian personality typology. *Journal of Abnormal & Social Psychology*, 68(1), 62-71.
- Stricker, L. J., & Ward, W. C. (2004). Stereotype threat, inquiring about test takers' ethnicity and gender, and standardized test performance. *Journal of Applied Social Psychology*, 34, 665-693.
- Tierney, J. P., Grossman, J. B., & Resch, N. L. (1995). *Making a difference: An impact study of Big Brothers/Big Sisters*. Philadelphia: Public/Private Ventures.
- Walpole, M. B., Burton, N. W., Kanyi, K., & Jackenthal, A. (2002). *Selecting successful graduate students: In-depth interviews with GRE users* (GRE Board Research Report No. 99-11R). Princeton, NJ: ETS.
- Walters, A. M., Lee, S., & Trapani, C. (2004). *Stereotype threat, the test-center environment, and performance on the GRE General Test* (GRE Board Research Report No. 01-03R). Princeton, NJ: ETS.
- Williams, J. M., & Nulty, D. D. (1986). Construct accessibility, depression and the emotional stroop task: Transient mood or stable structure? *Personality and Individual Differences*, 7(4), 485-491.

Witkin, H., & Goodenough, D. (1981). *Cognitive styles: Essence and origins: Field dependence and independence*. New York: International Universities Press.

Acknowledgements

Thanks to Ed Shea, William Monaghan, Richard Roberts, Walter Emmerich, Ralf Schulze, Jihyun Lee, and Lazar Stankov for contributions to and comments on this article.

R&D Connections is published by

ETS Research & Development
Educational Testing Service
Rosedale Road, 19-T
Princeton, NJ 08541-0001

Send comments about this publication to the above address or via the Web at:

<http://www.ets.org/research/contact.html>

Copyright © 2005 by Educational Testing Service. All rights reserved. Educational Testing Service is an Affirmative Action/Equal Opportunity Employer.

Educational Testing Service, ETS, the ETS logo, *Graduate Record Examinations*, and GRE are registered trademarks of Educational Testing Service. *The Praxis Series* is a trademark of Educational Testing Service.

SAT is a registered trademark of the College Entrance Examination Board.