Assessment Methods

Noncognitive characteristics can be assessed in many different ways, including self-assessments, interviews, and behavioral observations. The most commonly used methods shown in the table below include both well-established and recently developed assessment tools and are organized by source (self or other) and type (ratings or performance).

Self		Others	
Ratings	Performance	Ratings	Performance
Self-Assessments (Likert Type)	Situational Judgment Tests	Letters of Recommendation	Transcripts
Biodata		Others' Ratings (Likert Type)	Observations
Day Reconstruction Method			

Different assessments do not always give the same score on a trait. For example, self-ratings may not highly correlate with actual ability or test performance,[1] although the correlation can be increased by following certain procedures.

It also is not necessarily the case that one method is better than the other. Sometimes, two measures independently predict outcome criteria, each adding variance to the other. For example, it has been found that both self-reported and peer-rated Conscientiousness predicted school performance (controlling for intelligence) independent of one another.[2] **Note**: For a more detailed description of Conscientiousness and other personality variables, please see **Five-Factor Model**.

Faking is also an issue in noncognitive assessment, particularly when used in high-stakes applications. The fakeability factor of each assessment method is covered in the following descriptions, as well as psychometric topics related to the prediction of educational outcomes.

Self-Assessments

Self-assessments are the most widely used approaches for capturing students' noncognitive characteristics.

Most insights concerning the relationship between noncognitive qualities and educational or work-related outcomes stem from research conducted with questionnaires. Self-assessments usually ask individuals to describe themselves by answering a series of standardized questions. The answer format is usually a Likert-type rating scale (for example, in a range from Strongly Disagree to Strongly Agree) but other formats may also be used (such as Yes-No or open answer). Typically, questions assessing the same construct are aggregated and this aggregated score serves as an indicator of the relevant personality domain.

Self-assessments are a relatively easy, cost-effective, and efficient way of gathering information about the individual.[3] However, many issues need to be taken into account when developing a psychometrically sound questionnaire, such as number of points on a scale, scale point labels, and neutral point.[4] For instance, response scale format influences responses.[5] Respondents also vary in their use of the scale and, in general, there are large cultural effects in response style.[6] For example, young males tend to use extreme answer categories,[7] as do Hispanics.[8]

Respondents can also fake their responses to appear more attractive to a prospective employer or institution,[9, 10, 11] resulting in decreased validity.[12] ETS has conducted several mini-conferences addressing the faking problem and has identified several promising methods for collecting self-assessments, such as giving real-time warnings,[13] using a multidimensional forced-choice format (pitting equally attractive noncognitive factors such as "works hard" and "works well with others" against each other),[14] and using one's estimates of how others will respond to help control for faking.[15, 16] However, evidence for their effectiveness in controlling faking has not been demonstrated unequivocally.[17, 18]

Self-assessments have been used in many research studies involving noncognitive variables at the Center for Academic and Workforce Readiness and Success.

A selection of peer-reviewed articles and reports that ETS has published on the topic of self-assessments over the past few years follows:

Von Davier, M., Naemi, B., & Roberts, R. D. (in press). Factorial versus typological models for data from personality tests and survey scales. *Measurement*.

Preckel, F., Lipnevich, A. A., Boehme, K., Brandner, L., Georgi, K., Könen, T., Mursin, K., & Roberts, R. D. (in press). Morningness-eveningness and educational outcomes: The lark has an advantage over the owl at High School. *British Journal of Educational Psychology*.

MacCann, C., & Roberts, R. D. (in press). Just as smart but not as successful: Obese students obtain lower school grades but equivalent test scores to non-obese students. *International Journal of Obesity*.

MacCann, C., Fogarty, G., & Roberts, R. D. (2012). Strategies for success in vocational education: The role of time management. *Learning and Individual Differences*, *22*, 618–623.

Lipnevich, A. A., MacCann, C., Bertling, J. P., Naemi, B., & Roberts, R. D. (2012). Emotional reactions toward school situations: Relationships with academic outcomes. *Journal of Psychoeducational Assessment*, *30,* 387–401.

MacCann, C., Lipnevich, A. A., Burrus, J. B., & Roberts, R. D. (2012). The best years of our lives? Coping with stress predicts school grades, life satisfaction, and feelings about high school. *Learning and Individual Differences, 22*, 235–241.

Preckel, F., Lipnevich, A. A., Schneider, S., & Roberts, R. D. (2011). Chronotype, cognitive abilities, and academic achievement: A meta-analytic investigation. *Learning and Individual Differences*, *21*, 483–492.

Lipnevich, A. A., MacCann C., Krumm, S., Burrus, J., & Roberts, R. D. (2011). Mathematics attitudes and mathematics outcomes in U.S. and Belarusian middle school students. *Journal of Educational Psychology, 103,* 105–108.

MacCann, C., & Roberts, R. D. (2010). Development of a student health assessment system: Health knowledge, attitudes, and behaviors in middle-schoolers. *Educational Testing Service Research Report No: RR-10-04*. Princeton, NJ: Educational Testing Service.

MacCann, C., Duckworth, A., & Roberts, R. D. (2009). Identifying the major facets of Conscientiousness in high school students and their relationships with valued educational outcomes. *Learning and Individual Differences*, *19*, 451–458.

Liu, O. L., Rijmen, F., MacCann, C., & Roberts, R. D. (2009). The assessment of time management in middle school students. *Personality and Individual Differences, 47*, 174–179.

Situational Judgment Test (SJT)

A situational judgment test (SJT) is one in which participants are asked how best to or how they might typically deal with some kind of situation. For example, a situation might be a group project in which one member did not help out. Possible responses might be to:

- talk to the nonparticipating member in private
- talk to the nonparticipating member in front of the group
- let the incident pass without comment

Situations can be described in words or videotaped, and responses can include multiple choice, constructed

response, and ratings ("How good would this response be?").[19] SJTs can be regarded as fairly simple, economical simulations of job tasks.[20]

SJTs have been shown to predict many different criteria such as college success,[21, 22] army leadership,[23, 24] and managerial performance.[25] Though applications in education have been relatively limited, applying SJTs as a predictor in educational domains has received increased interest.[26, 27]

SJTs may be developed to reflect more subtle and complex judgment processes than are possible with conventional tests. The methodology of SJTs enables the measurement of many relevant attributes of individuals, including leadership, the ability to work with others, achievement orientation, self-reliance, dependability, sociability, agreeableness, social perceptiveness, and conscientiousness.[28, 29] Numerous SJTs, ranging from print-based measures of business analysis and problem solving[30] to video-based measures of communication skills,[31] have been developed.

Research on SJTs has revealed that respondents are able to improve their score in a retest[32] or after coaching,[33] although the improvement may be small. SJTs appear to be less susceptible to faking compared to Likert-type self-assessments, where the improvement due to incentives can be up to a full standard deviation.

Among the following are peer-reviewed articles, reports, and book chapters that ETS has published on the topic of SJTs:

Burrus, J., Betancourt, A., Holtzman, S., Minsky, J., MacCann, C., & Roberts, R. D. (2012). Emotional intelligence relates to well-being: Evidence from the Situational Judgment Test of Emotional Management. *Applied Psychology: Health and Well-Being, 4,* 151–166.

MacCann, C., Fogarty, G. J., Zeidner, M., & Roberts, R. D. (2011). Coping mediates the relationship between emotional intelligence (EI) and academic achievement. *Contemporary Educational Psychology*, *36*, 60–70.

MacCann, C., Wang, P., Matthews, G., & Roberts, R. D. (2010). Examining self-report versus other reports in a situational judgment test of emotional abilities. *Journal for Research in Personality, 44*, 673–676.

Wang, L., MacCann, C., Zhuang, X., Liu, L., & Roberts, R. D. (2009). Assessing teamwork and collaboration in high school students: A multimethod approach. *Canadian Journal of School Psychology, 24,* 108–124.

MacCann, C., & Roberts, R. D. (2008). New paradigms for assessing emotional intelligence: Theory and data. *Emotion*, *8*, 540–551.

Zhuang, X., MacCann, C., Wang, L., Liu, L., & Roberts, R. D. (2008). Development and validity evidence supporting a teamwork and collaboration assessment for high school students. *Educational Testing Service Research Report No: RR-08-50*. Princeton, NJ: ETS.

Biodata

Biographical data or "biodata" have been explored for college admissions use in the United States[34] and Chile.[35] Biodata are typically obtained by asking standardized questions about individuals' past behaviors, activities, or experiences. A sample question could be: "How often in the last two weeks have you eaten fast food?" Respondents are given multiple-choice answer options or are requested to answer in an open format (e.g., frequency).

ETS collects biodata in most of our research studies, including:

- Gender
- Race/Ethnicity
- Age (specific via open response or a range via multiple choice)
- Socio-economic Status (via items that include a range of income, financial aid status, number of books in a home, etc.)
- Grade Point Average (GPA) (via open response and restricted ranges) & Coursework (courses taken, grades, AP® courses, etc)
- SAT®/ACT® Scores
- Citizenship Status
- College Enrollment Status

- College Major
- Attendance
- Languages (first language, language spoken at home, etc)
- Parental Status
- Access to Computers
- Work Status (full-time, part-time, unemployed, hours worked)
- Parental Education
- Level of Schooling (current and aspirations)
- Health Information (weight, height)

We have developed a biodata (documented accomplishments) measure that produced scores for six scales:[36, 37]

- Academic Achievement
- Leadership
- Practical Language
- Aesthetic Expression
- Science
- Mechanical

For the Leadership category, for example, items were:

- *Was on a student-faculty committee in college. Yes/No*
- *If YES: Position, organization and school?*

Measures of biodata have been found to be incrementally valid beyond the SAT test and the Big Five in predicting students' performance in college.[38] Obviously, biodata can be faked but there are several ways to minimize faking.[39, 40] Asking students to verify with details, for example, can minimize faking.

Day Reconstruction Method (DRM)

The day reconstruction method (DRM) measures how people spend their time and how positively or negatively they feel while conducting certain activities.[41] Participants list details from several episodes they participated in during the previous day and state when each episode occurred, what they were doing, who they were with, and how they felt. Research has demonstrated that recalling episodes from the previous day is close enough in time for people to accurately recall emotions without bias.[42]

In our initiative, we have used the DRM in two ways. First, we have used it as a measure of hedonic well-being or the daily experience of positive and negative emotions. For example, we have found that hedonic well-being, is correlated with emotional intelligence.[43] We have also used the DRM as part of a time management intervention, and found evidence that suggests it may be useful in improving this skill.[44] We believe the act of recording the events of one's day in detail can lead to improved self-awareness of how time is spent.

Day Reconstruction:

- **Lawrenceville Studies:** These studies specifically focused on improving the time management skills of high school students at a private school; the DRM was used as a homework assignment in an effort to increase awareness of how time was spent and how one felt during each activity for each day. A paper currently under peer review has been written on this topic and is available upon request.
- **Army Research Institute (ARI) studies:** The focus of the ARI studies was to assess emotional abilities (Phase I), and their impact on well-being (Phase II), stress (Phase III) and leadership (Phase IV). Two- and four-year college students responded to the DRM during Phase II to assess the time use and emotions associated with activities during the previous day.[45]

Other-Ratings

Other-ratings are assessments in which others (e.g., supervisors, trainers, colleagues, friends, faculty advisors, coaches) rate individuals on various noncognitive qualities. This method has a long history and countless studies have been conducted that employed this method to gather information.[46] Other-ratings have an advantage over self-ratings in that they preclude socially desirable responding, although they do permit rating biases. Self- and other-ratings do not always agree,[47] but other-ratings are often more predictive of outcomes than self-ratings.[48, 49]

Other-Ratings Used at ETS:

- *ETS*® Personal Potential Index/ETS Summer Intern Standardized Letter of Recommendation Study: In this study, the advisors and faculty members of prospective graduate students rated these students

in the following areas: Knowledge and Creativity, Communication Skills, Teamwork, Resilience, Planning and Organization, and Ethics and Integrity on a scale that ranged between Below Average and Truly Exceptional.[50]

- **Studies of other-reported noncognitive factors.** Because of the validity threat posed by faking in self-assessments, many of the papers described in that section also include other-reports (i.e., some combination of teacher-, parent- and/or peer-reports). A recent study also explored the topic of other-reported SJTs, to present knowledge the first study of its kind.[51]

Transcripts

Transcripts contain information on the courses students have taken, earned credits, grades, and grade point average. As official records, transcript information can be taken as more accurate than self-reports. Transcript data can be standardized and used in validity studies. For example, the U.S. National Center for Educational Statistics supports an ongoing collection of transcripts based on the **National Assessment of Educational Progress (NAEP)** High School Transcript Study, which classifies courses, computes grade point average, and links resulting data to NAEP achievement scores.

Interviews

Interviews are the most frequently used method of personnel selection,[52] but also are used for school admissions, promotions, scholarships, and other awards. Interviews vary in their content and structure. In a structured interview, questions are prepared before the interview starts. An unstructured interview simply represents a free conversation between an interviewer and interviewee, giving the interviewer the freedom to adaptively or intuitively switch topics. Research has shown that unstructured interviews lack predictive validity[53] or show lower predictive validity than structured interviews.[54]

The best practices for conducting interviews are:[55]

- High degree of structure
- Selection of questions according to job requirements

- Assessment of aspects that cannot be better assessed with other methods
- Scoring with pre-tested, behavior-anchored rating scales
- Empirical examination of each question
- Rating only after the interview
- Standardized scoring
- Training of interviewers

Structured interviews can be divided into three types:

- Behavioral description interview — involves questions that refer to past behavior in real situations, also referred to as job-related interview[56]
- Situational interview — uses questions that require interviewees to imagine hypothetical situations (derived from critical incidents) and state how they would act in such situations[57]
- Multimodal interview — combines the two approaches above and adds unstructured parts to ensure high respondent acceptance[58]

Analyses of predictive validity of interviews for job performance[59, 60, 61, 62] have shown that they are good predictors of job performance, add incremental validity above and beyond general mental ability, and that behavioral description interviews show a higher validity than situational interviews. Interviews are less predictive of academic performance as compared to job-related outcomes. Predictive validity probably also depends on the content of the interview, but the analyses aggregated interviews with different contents.

Behavioral Observations

Behavioral observations entail watching observable activities of individuals and keeping records of the relevant activities.[63] Records can vary from videos, photographs, and cassette recordings to notes taken by the observer. The general assumption behind this method is that individuals vary in observable behaviors; this variation is stable over time and across different situations, and can be regarded as an indicator of a personality trait.[64]

One form of behavioral observation often used in selection is the assessment center. Assessment centers can comprise many different methods (including achievement tests), but they feature role play and presentation

tasks. In these tasks, participants are asked to act in a simulated situation. These situations are designed in such a way that a certain behavior can be highlighted (e.g., assertiveness). Analysis has shown that assessment centers moderately predict job and training performance but do not add incremental validity beyond general mental ability.[65]

A strength of assessment centers for measuring personality is that they are performance-based rather than opinion-based self-assessments. As such, they are less easily faked than are self-assessments. On the other hand, a drawback to assessment centers is that they assess maximum performance, which may not be representative of typical behavior.

Letters of Recommendation

Letters of recommendation can be seen as a more subjective form of other-ratings and have been extensively used in a broad range of situations.[66] They provide stakeholders with detailed information about the applicants' past performance, with the writer's opinion about the applicant being expressed in the form of an essay.

In response to a major drawback of letters of recommendation — their nonstandardized format — a more structured system, initially coined the Standardized Letter of Recommendation[67] and now the *ETS® Personal Potential Index* (ETS® PPI),[68] has been developed. This assessment system prompts faculty members to respond to specific items using a Likert scale, in addition to eliciting comments. It has been used operationally at ETS for selecting summer interns and fellows,[69] [70] as well as through Project 1000 for the selection of graduate student applicants.[71] ETS PPI has been offered as a supplement to *GRE® tests* beginning in 2009.[72]

Footnotes

[1] Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67,* 280–296.

[2] Bratko, D., Chamorro-Premuzic, T., & Saks, Z. (2006). Personality and school performance: Incremental validity of self- and peer-ratings over intelligence. *Personality and Individual Differences, 41,* 131–142.

[3] Zeidner, M., Matthews, G., & Roberts, R. D. (2009). *What we know about emotional intelligence: How it affects learning, work, relationships, and our mental health.* Cambridge, MA: MIT Press.

[4] Krosnick, J. A., Judd, C. M., & Wittenbrink, B. (2005). Attitude measurement. In D. Albarracin, B. T. Johnson, & M. P. Zanna (Eds.), *Handbook of attitudes and attitude change.* Mahway, NJ: Erlbaum.

[5] Rammstedt, B., & Krebs, D. (2007). Does response scale format affect the answering of personality scales? Assessing the big five dimensions of personality with different response scales in a dependent sample. *European Journal of Psychological Assessment, 23,* 32–38.

[6] Harzing, A-W. (2006). Response styles in cross-national survey research. *International Journal of Cross-cultural Management, 6(2),* 243–266.

[7] Austin, E. J., Deary, I. J., & Egan, V. (2006), Individual differences in response scale use: Mixed Rasch modeling of response to NEO-FFI items. *Personality and Individual Differences, 40,* 1235–1245.

[8] Marin, G. Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics. *Journal of Cross-Cultural Psychology, 23(4),* 498–509.

[9] Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36,* 341–357.

[10] Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59,* 197–210.

[11] Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods, 7,* 168–190.

[12] Pauls, C. A., & Crost, N. W. (2005). Effects of different instructional sets on the construct validity of the NEO-PI-R. *Personality and Individual Differences, 39,* 297–308.

[13] Sackett, P. R. (2006). *Faking and coaching effects on non-cognitive predictors.* Paper presented at the ETS Mini-conference on Faking in Noncognitive Asssessments. Princeton, NJ: Educational Testing Service.

[14] Stark, S., Chernyshenko, O. S., & Drasgow, F. (2005). An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: An application to the problem of faking in personality assessment. *Applied Psychological Measurement, 29,* 184–201.

[15] Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science, 306,* 462–466.

[16] Prelec, D. & Weaver, R. G. (2006). Truthful answers are surprisingly common: Experimental tests of Bayesian truth serum. Paper presented at the ETS Mini-conference on Faking in Noncognitive Assessments. Princeton, NJ: Educational Testing Service.

[17] Converse, P. D., Oswald, F. L., Imus, A., Hedricks, C., Roy, R. & Butera, H. (2008). Comparing personality test formats and warnings: Effects on criterion-related validity and test-taker reactions. *International Journal of Selection and Assessment, 16 (2),* 155–169.

[18] Heggestad, E. D., Morrison, M., Reeve, C. L., & McCloy, R. A. (2006). Forced-choice assessments for selection: Evaluating issues of normative assessment and faking resistance. *Journal of Applied Psychology, 91,* 9–24.

[19] McDaniel, M. A., Morgesen, 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.

[20] Kyllonen, P. C., & Lee, S. (2005). Assessing problem solving in context. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence,* 11–25. Thousand Oaks, CA: Sage.

[21] Lievens, F., & Coestsier, P. (2002). Situational tests in student selection: An examination of predictive validity, adverse impact, and construct validity. *International Journal of Selection and Assessment, 10,* 245–257.

[22] Oswald et al. (2004).

[23] Krokos, K. J., Meade, A. W., Cantwell, A. R., Pond, S. B., & Wilson, M. A. (2004). *Empirical keying of situational judgment tests: Rationale and some examples.* Paper presented at the annual conference of the Society for Industrial and Organizational Psychology, Chicago, IL.

[24] Legree, P. J. (1995). Evidence for an oblique social intelligence factor. *Intelligence, 21*, 247–266.

[25] Howard, A., & Choi, M. (2000). How do you assess a manager's decision-making abilities? The use of situational inventories. *International Journal of Selection and Assessment, 8*, 85–88.

[26] Lievens, F., Buyse, T., & Sackett, P. R. (2005a). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90*, 442–452.

[27] Oswald et al. (2004).

[28] Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89*, 187–207.

[29] Waugh, G. W., & Russell, T. L. (2003). *Scoring both judgment and personality in a situational judgment test*. Paper presented at the 45th Annual Conference of the International Military Testing Association, Pensacola, FL.

[30] Kyllonen & Lee (2005).

[31] 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.

[32] Lievens, F., Buyse, T., & Sackett, P. R. (2005b). Retest effects in operational selection settings. Development and test of a framework. *Personnel Psychology, 58*, 981–1007.

[33] Cullen, M. J., Sackett, P. R., & Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment, 14*, 142–155.

[34] Oswald et al. (2004).

[35] Delgalarrando, M. G. (July 9, 2008). *Validan plan de admisión complementaria a la UC (p. 9)*, El Mercurio, Santiago, Chile.

[36] Baird, L. L., & Knapp, J. E. (1981). *The inventory of documented accomplishments for graduate admissions: Results of a field trial study and its reliability, short-term correlates, and evaluation*. (ETS Research Rep. No. 81-18, *GRE®* Board Research Rep. No. 78-3R). Princeton, NJ: Educational Testing Service.

[37] Stricker, L. J., Rock, D. A., & Bennett, R. E. (2001). Sex and ethnic-group differences on accomplishment measures. *Applied Measurement in Education, 14,* 205–218. [G]

[38] Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology, 89*, 187–207.

[39] Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*, 1–23.

[40] Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M. A., & Ramsay, L. J. (2003). Impact of elaboration on socially desirable responding and the validity of biodata measures. *Journal of Applied Psychology, 88*, 979–988.

[41] Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The Day Reconstruction Method. *Science, 306,* 1776–1780.

[42] Kahneman et al. (2004).

[43] Burrus, J., Holtzman, S., Minsky, J., MacCann, C., & Roberts, R. D. (2012). Emotional intelligence relates to well-being: Evidence from the situational judgment test of emotional management. *Applied Psychology: Health and Well-Being, 4,* 151–166.

[44] Burrus, J., Jackson, T., Holtzman, S., Roberts, R. D., & Mandigo, T. (2012). Examining the efficacy of a time management intervention for high school students. In review at the *Journal of School Psychology.*

[45] Burrus, J., Holtzman, S., Minsky, J., MacCann, C., & Roberts, R. D. (2012). Emotional intelligence relates to well-being: Evidence from the situational judgment test of emotional management. *Applied Psychology: Health and Well-Being, 4,* 151–166.

[46] Tupes, E. C., & Christal, R. E. (1961/1992). Recurrent personality factors based on trait ratings. *Journal of Personality, 60*, 225–251.

[47] Oltmanns, T. F., & Terkheimer, E. (2006). Perceptions of self and others regarding pathological personality traits. In R. Krueger & J. Tackett (Eds.), Personality and psychopathology: Building bridges. New York: Guilford.

[48] Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

[49] Wagerman, S., & Funder, D. C. (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. *Journal of Research in Personality, 41*, 221–229.

[50] Walters, A. M., Kyllonen, P. C., & Plante, J. W. (2006). Developing a Standardized Letter of Recommendation. *Journal of College Admission, 191,* 8–17.

[51] MacCann, C., Wang, P., Matthews, G., & Roberts, R. D. (2010). Examining self-report versus other reports in a situational judgment test of emotional abilities. *Journal for Research in Personality, 44*, 673–676.

[52] Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. *Personnel Psychology, 52*, 359–391.

[53] Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology, 35*, 281–322.

[54] Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology. *Psychological Bulletin, 124*, 262–274.

[55] Schuler, H. (2002). *Das Einstellungsinterview*. Göttingen: Hogrefe.

[56] Janz, T., Hellervik, L., & Gilmore, D. C. (1986). Behavior Description Interviewing (BDI). Boston: Allyn & Bacon.

[57] Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview (SI). *Journal of Applied Psychology, 65*, 422–427.

[58] Schuler, H. (2002). *Das Einstellungsinterview (Multimodales Interview [MMI])*. Göttingen: Hogrefe.

[59] Huffcutt, A. I., Conway, J. M., Roth, P. L., & Klehe, U. C. (2004). The impact of job complexity and study design on situational and behavior description interview validity. *International Journal of Selection and Assessment, 12*, 262–273.

[60] Marchese, M. C., & Muchinski, P. M. (1993). The validity of the employment interview: A meta-analysis. *International Journal of Selection and Assessment, 1*, 18–26.

[61] McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D., (1994). The validity of the employment interview: A meta-analysis. *International Journal of Selection and Assessment, 79,* 599–616.

[62] Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology. *Psychological Bulletin, 124*, 262–274.

[63] Cohen, R. J., & Swerdlik, M. E. (2005). *Psychological testing and assessment (6th Ed.)*. New York: McGraw-Hill.

[64] Stemmler, G. (2005). Studierendenauswahl durch Hochschulen: Ungewisser Nutzen. *Psychologische Rundschau, 56*, 125–127.

[65] Schmidt & Hunter (1998).

[66] Arvey, R. D. (1979). Unfair discrimination in the employment interview: Legal and psychological aspects. *Psychological Bulletin, 86*, 736–765.

[67] Walters, A. M., Kyllonen, P. C., & Plante, J. W. (2003). *Preliminary Research to Develop a Standardized Letter of Recommendation*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA. (2006). Developing a standardized letter of recommendation *The Journal of College Admission, 191*, 817.

[68] ETS (2009). *ETS®* Personal Potential Index.

[69] Kyllonen, P. C., & Kim, S. (2004). *Personal qualities in Higher Education: Dimensionality of faculty ratings of graduate school applicants*. Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

[70] Kim, S., & Kyllonen, P. C. (2008). Rasch measurement in developing faculty ratings of students applying to graduate school. *Journal of Applied Measurement, 9(2)*, 168–181.

[71] Liu, O. L., Minsky, J., Ling, G., & Kyllonen, P. (2007). *The Standardized Letter of Recommendation: Implications for Selection*. ETS Research Report RR-07-38. Princeton, NJ: ETS.

[72] See ETS (2009). *ETS*® Personal Potential Index. Kyllonen, P. C. (2008). *The research behind the ETS® Personal Potential Index*. Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Products/PPI/10411_PPI_bkgrd_report_RD4.pdf.

ETS

Listening. Learning. Leading.®

www.ets.org