Assessing practical intelligence in business school admissions: A supplement to the graduate management admissions test

Jennifer Hedlund a,*, Jeanne M. Wilt b, Kristina L. Nebel b, Susan J. Ashford b, Robert J. Sternberg c

a Department of Criminology, Central Connecticut State University, 1615 Stanley Street, New Britain, CT 06050, United States
b University of Michigan Business School, United States
c Yale University, United States

Received 25 February 2005; received in revised form 17 July 2005; accepted 20 July 2005

Abstract

The Graduate Management Admission Test (GMAT) is the most widely used measure of managerial potential in MBA admissions. GMAT scores, although predictive of grades in business school, leave much of the variance in graduate school performance unexplained. The GMAT also produces disparities in test scores between groups, generating the potential for adverse impact in the admissions process. We sought to compensate for these limitations by adding measures of practical intelligence to the admissions process in an MBA program. We developed two approaches to measuring practical intelligence, one knowledge-based and the other skill-based. We administered the resulting measures to two samples of incoming MBA students (total N=792). Across the two studies, we found that scores on both measures predicted success inside and outside the classroom and provided small, yet significant, increments beyond GMAT scores and undergraduate GPA in the prediction of variance in MBA performance. We further found that these measures exhibited less disparity across gender and racial/ethnic groups than did the GMAT. These findings, although preliminary, suggest the potential value of considering a broader range of abilities in admissions testing.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Practical intelligence; Admissions testing; Leadership
1. Introduction

The admission process in Masters of Business Administration (MBA) programs involves assessing each candidate’s demonstrated and potential abilities to be a successful student and business leader. The Graduate Management Admissions Test (GMAT) is perhaps the most widely used uniform criterion in business school admissions (Dobson, Krapljan-Barr, & Vielba, 1999; Hancock, 1999; Wright & Palmer, 1994). Like many standardized admission tests, the GMAT consistently explains about 20% of the variance in graduate GPA (Ahmadi, Raiszadeh, & Helms, 1997; Graham, 1991; Hancock, 1999; Nilsson, 1995; Paolillo, 1982; Wright & Palmer, 1994; Youngblood & Martin, 1982). Although 20% is impressive in regard to psychological measurement, 80% of the variance in school performance remains unexplained. Furthermore, there is evidence to suggest that the GMAT’s predictive validity may be limited largely to performance in graduate school (Bretz, 1989).

Another potential limitation with the use of standardized admissions tests is their potential for adverse impact (Bracey, 2001; Darling-Hammond, 1991). Researchers have consistently found mean differences on tests of general cognitive ability and related abilities for different racial/ethnic groups; most notably, blacks tend to score one standard deviation lower than whites on these tests (Hartigan & Wigdor, 1989; Jensen, 1998; Schmidt, 1988; Williams, 2000). The GMAT, in particular, has been found to exhibit disparities in test scores across both gender and racial/ethnic subgroups (Dobson et al., 1999; Hancock, 1999). These differences tend to favor males over females and whites over blacks. Most notably, blacks tend to score more than one standard deviation lower than whites on the GMAT. Identifying a supplement to the GMAT that could potentially reduce, or compensate for, these differences while at the same time increasing prediction, could be of great benefit to business school admissions.

The purpose of our research was to explore alternative ways to assess a business school candidate’s potential for success that address, although certainly do not fully circumvent, some of the limitations identified above. The goal is not to replace the GMAT or comparable assessments, but rather to supplement them and thus to improve the process of business school admissions. Our efforts focused on the incorporation of measures of practical abilities as complements to the analytically oriented GMAT, with the aim of improving the prediction of business-school performance and, ultimately, business success. This article presents the development and preliminary validation of measures of practical intelligence for use in MBA admissions.

2. Exploring alternatives to conventional standardized admissions tests

Before pursuing a supplement to the GMAT, it was important to consider what constructs the test purports to measure, and thus, what constructs it does not intend to assess. The GMAT is comparable to other standardized admissions test (e.g., SAT, LSAT) in that it consists of multiple-choice questions that measure verbal and quantitative skills. The GMAT further includes an essay component aimed at measuring analytical skills. According to Jaffe and Hilbert (1994), “The purpose of the GMAT is to measure your ability to think systematically and to employ the verbal and mathematical skills that you have acquired throughout your years of schooling” (p. 3). Thus, like the SAT (Frey & Detterman, 2004), the GMAT can be characterized as a traditional measure of intelligence, or a test of general cognitive ability (g).
Given the above-mentioned limitations associated with standardized admissions tests, researchers have suggested broadening the attributes considered in selection and admissions decisions (Sacket, Schmitt, Ellington, & Kabin, 2001). These attributes might include personality, motivation, interpersonal skills, and prior experience, to name a few. Some researchers argue that measures of these attributes, such as biographical data, structured interviews, personality tests, and situational-judgment tests, are potentially useful additions to selection and admission decisions because they tap “noncognitive” factors relevant to performance (Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Sacket et al., 2001). Furthermore, there is evidence to suggest that such measures reduce subgroup differences and increase predictive validity (Hough, Oswald, & Ployhart, 2001; Pulakos & Schmitt, 1996).

These so-called noncognitive predictors have traditionally been used in business school admissions, along with GMAT scores. Applicants often are asked to provide resumes, letters of recommendation, and essays, which are used to evaluate work experience as well as academic and leadership potential. Some programs further recommend or require an interview, which can be used to evaluate interpersonal skills, in addition to the previously mentioned characteristics. Unfortunately, most of these items are difficult to quantify and receive limited weight in admissions decisions. In reality, they typically are used only for a small number of candidates who have GMAT scores that fall near or below an established cutoff score. If these attributes are important for success as a business school student and as a business leader, ideally, they should be considered along with GMAT scores for all candidates.

Oswald et al. (2004) explored the potential value of adding quantifiable measures of noncognitive attributes to college admissions. Specifically, they developed a biographical measure and a situational-judgment inventory for college undergraduate students. Biographical (biodata) measures consisted of standardized questions regarding an individual’s prior experiences. Situational-judgment inventories consisted of hypothetical problems relevant to those encountered in the performance domain. These measures were developed to assess 12 dimensions of college performance, such as learning, leadership, citizenship, adaptability, perseverance and ethics.

Oswald et al. (2004) examined the validity of these measures relative to the SAT/ACT in predicting school performance. They found that scores on 12 biodata scales and an overall score on the SJI accounted for significant variance in GPA, absenteeism, and self- and peer-ratings of college performance beyond SAT/ACT scores and the Big Five personality traits. These increments ranged from $\Delta R^2 = .06$ with GPA to $\Delta R^2 = .22$ with self-ratings of performance. In regard to group differences, they found that females scored lower than males on the SAT/ACT, but tended to score higher than males on the biodata items and the SJI. They also found that subgroup differences on the biodata items tended to be relatively small compared to those observed for the SAT/ACT. Additionally, there were no significant differences across subgroups on the SJI. These findings suggest that inclusion of noncognitive measures may help reduce the potential for adverse impact, while also increasing prediction in comparison with relying solely on traditional admissions tests.

An alternative but related strategy for dealing with the limitations of standardized admissions tests is based on a different conceptualization of cognitive abilities. In particular, some researchers propose that definitions of intelligence should include aspects such as interpersonal intelligence (Gardner, 1993, 1999), emotional intelligence (Goleman, 1998; Mayer, Salovey, & Caruso, 2000), and creative and practical intelligence (Sternberg, 1985, 1997, 1999; Sternberg et al., 2000). Rather than view all alternatives to general cognitive ability tests as noncognitive, these perspectives view cognitive ability tests as measuring one aspect of intelligence. These broader conceptualizations of intelligence recognize
that individuals have different strengths, and that these strengths may not be identified through
traditional approaches to measuring intelligence.

Practical intelligence is a component of a broader definition of intelligence, which also includes
creative and analytical intelligence (Sternberg, 1997, 1999). Practical intelligence is defined as the
ability that individuals use to find a more optimal fit between themselves and the demands of
the environment through adapting to the environment, shaping (or changing) the environment, or selecting
a new environment in the pursuit of personally valued goals (Sternberg, 1985, 1997). It can be
characterized as “street smarts” or “common sense,” and can be contrasted with analytical intelligence
or “book smarts.” In order to succeed in life, individuals need “common sense” as well as “book
smarts.” The concept of practical intelligence encompasses many of the attributes that others have
labeled as noncognitive, such as interpersonal skills, perseverance, and good judgment. These attributes
are important to performance, whether it be in school or on-the-job, but are not necessarily captured by
standardized admissions or selection tests. We argue that measures of practical intelligence offer one
means of addressing the limitations associated with more analytically oriented admissions tests, in
particular, the GMAT.

3. Practical intelligence and MBA admissions

Measures of so-called practical abilities have been used in the area of job selection for nearly 50 years.
For example, in-basket tests were designed to assess an individual’s ability to deal with job-related tasks
under some of the same constraints (e.g., deadlines) found on the actual job (Frederiksen, 1966;
Frederiksen, Saunders, & Wand, 1957). Assessment centers also are used to observe an individual’s
performance in situations that have been created to represent aspects of the actual job situation.
Assessment centers typically present small groups of individuals with a variety of tasks, including in-
basket tests, simulated interviews, and simulated group discussions (Bray, 1982; Thornton & Byham,
1982). Responses to these simulations are considered to represent the actual, or close approximations of,
responses that individuals would exhibit in real situations. Situational-judgment inventories represent yet
another approach to measuring practical abilities, and have even been viewed as low-fidelity simulations
of actual work situations (Motowidlo, Dunnette, & Carter, 1990). There is also evidence to suggest that
some of these measures (e.g., work sample tests) can exhibit higher validity than general cognitive
ability tests (Schmidt & Hunter, 1998). In other words, there is evidence to suggest that measures
intended to reflect more closely actual job performance may be equally, if not more, valid than measures
of general cognitive ability.

We chose to explore the use of measures of practical abilities as supplements to the GMAT in MBA
admissions. In particular, we pursued two alternative methods of measuring practical intelligence, one
knowledge-based and the other skill-based. These measures draw upon prior efforts to assess practical
abilities in the area of job performance.

3.1. Tacit knowledge

Practical intelligence often has been equated with the notion of “common sense.” It involves
knowing how to navigate effectively through the problems of everyday life. Individuals who are
successful at solving these everyday problems also are said to rely, to some extent, on their “intuition.”
In other words, they develop effective solutions to problems without necessarily being able to explain or justify their decisions. This “intuition” or “common sense” has been attributed in the practical-intelligence literature to tacit knowledge (see Polanyi, 1976; Sternberg, Wagner, Williams, & Horvath, 1995). The concept of tacit knowledge reflects the idea that much of the knowledge relevant to performance is acquired through everyday experiences without conscious intent and guides our actions without being easily articulated.

Sternberg and his colleagues (Sternberg & Horvath, 1999; Sternberg et al., 2000) have focused on tacit knowledge as a means of providing insight into practical intelligence. They have studied tacit knowledge in domains as diverse as bank management, sales, academic psychology, primary education, clerical work, and military leadership (Hedlund et al., 2003; Sternberg & Wagner, 1993; Sternberg et al., 2000; Wagner, 1987; Wagner & Sternberg, 1985; Wagner et al., 1999).

Tacit knowledge (TK) typically is measured via situational-judgment inventories (SJIs). Individuals are presented with written descriptions of situations that represent actual situations or approximations of actual situations in the domain of interest (e.g., a salesperson making a phone solicitation). They are asked to rate the quality or appropriateness of potential solutions. Then their responses are scored relative to a standard based on expert or consensus judgment. This score is considered to be a measure of an individual’s tacit knowledge and is viewed as an indicator of his or her practical intelligence.

Previous research has examined the relationship of TK scores to domain-specific experience, general cognitive ability, and various indicators of performance. Generally, individuals with greater experience in a domain (e.g., business managers vs. business students) receive better TK scores (Hedlund, Sternberg, & Psotka, 2001; Sternberg, Wagner, & Okagaki, 1993; Wagner, 1987). Tacit knowledge scores also correlate fairly consistently with performance across a variety of domains. Individuals with higher TK scores have been found to have higher salaries, better performance ratings, more productivity, and to work in more prestigious institutions (Hedlund et al., 2003; Sternberg et al., 1993, 1995; Wagner, 1987; Wagner & Sternberg, 1985).

Finally, TK tests appear to tap abilities that are distinct from those measured by traditional intelligence or ability tests. The correlations between scores on TK tests and scores on traditional intelligence tests have ranged from negative to moderately positive (Sternberg et al., 2000, 2001, 1993; Wagner & Sternberg, 1985). More importantly, TK scores have been found to explain performance above and beyond that accounted for by tests of general cognitive ability (Hedlund et al., 2003; Wagner & Sternberg, 1990). Colonia-Willner (1998) also found that TK significantly predicted an index of managerial skill, whereas psychometric and verbal reasoning did not.

Other researchers have demonstrated the potential value of using SJIs to measure practical abilities (e.g., Chan & Schmitt, 1998; Fox & Spector, 2000; Pulakos, Schmitt, & Chan, 1996). Fox and Spector administered a SJI to undergraduate students participating in a simulated interview. The students were asked to select the response they would most likely or least likely take to several work-related situations. They found that practical intelligence significantly predicted evaluations of the interviewee’s qualifications. They also found that scores on the practical-intelligence test exhibited a moderate, significant correlation (.25) with a measure of general intelligence. Pulakos et al. (1996), using a SJI specifically designed for entry-level professionals in a federal investigative agency, found that practical intelligence predicted both peer and supervisory ratings of performance, and that the effects of practical intelligence were not accounted for by a test of general cognitive ability. Finally, Chan and Schmitt (1998) reported that SJIs tend to correlate with performance ratings for various jobs in the range of
3.2. Practical problem-solving skills

Another approach to assessing practical intelligence focuses on the skills involved in solving practical problems. The types of problems individuals encounter in their everyday lives differ from those found on traditional admissions tests, in that they tend to be poorly defined, to lack complete information, to have multiple possible solutions, and to be embedded in a context (Neisser, 1976; Sternberg, 1997; Wagner & Sternberg, 1986). According to Sternberg (1985, 1997), individuals who effectively solve practical problems are able to recognize that a problem exists, to define the problem clearly, to allocate appropriate resources to the problem, to formulate strategies for solving the problem, to monitor their solutions, and to evaluate the outcomes of those solutions. Furthermore, in order to understand the problem in the first place, individuals need to be able to filter relevant information from irrelevant information, relate new information to existing knowledge, and compile information into a meaningful picture. The effective use of these skills to solve practical, everyday problems can be viewed as an indicator of one’s practical intelligence. Thus, measures of practical problem-solving skills can serve as an alternative to knowledge-based measures of practical intelligence.

4. Study overview

The purpose of our research was twofold: (1) to develop and compare two alternative measures of practical intelligence, and (2) to validate those measures as potential supplements to the GMAT. In regard to the first goal, we chose to explore a knowledge-based and a skill-based approach to measuring practical intelligence. Although TK tests have exhibited concurrent validity with performance in academic and work settings, it is not known how well they will predict future performance. In other words, it is not clear if they are suitable for applicants who are not required to possess domain-specific knowledge (e.g., business school applicants). Therefore, we also considered using a measure of an applicant’s ability (or potential) to solve problems like those that might someday be encountered in the business world. In regard to the second goal, we evaluated the new measures on four factors: (a) predictive validity relative to indicators of academic and employment success; (b) incremental validity relative to GMAT scores and undergraduate GPA; (c) potential for adverse impact via disparities in test scores; and (d) applicants’ reactions to the measures (e.g., face validity, endorsement). We first describe the development of the practical intelligence (PI) measures. Then we present two studies aimed at assessing their validity.

5. Instrument development

The development of both PI measures began with the identification of the types of problems MBA graduates might experience as business leaders. To identify such problems, we considered existing taxonomies of managerial/leadership skills (see Mintzberg, 1980; Yukl, 1998), along with the types of business functions (marketing, accounting, information systems, general management) and business settings (finance, manufacturing, technology, nonprofit) in which MBA graduates typically work. We
generated 12 problem ideas that incorporated at least one leadership skill (e.g., managing conflict), one business function (e.g., general management), and one business setting (e.g., technology). In consultation with business-school faculty and admissions personnel, we selected 6 of the most promising ideas for further measurement development.

5.1. Measures of problem solving skills

For the skill-based measure of practical intelligence, we sought to create items with enough detail to allow an individual to demonstrate various problem-solving skills, such as problem recognition, solution generation, and outcome monitoring. We drew upon several techniques that have been used to measure practical abilities in the area of personnel selection (e.g., in-basket tests, work samples, case-based interviews). Because the format of our measures was unique to our purpose, we labeled them case scenario problems (CSPs). The CSPs presented a fictitious business case, which consisted of a brief overview of the problem, the respondent’s role, and a history of the organization. In addition, respondents were provided various documents such as organizational charts, departmental memos, e-mail correspondence, financial tables, and/or product descriptions. In developing the materials, we sought to include enough information to assess problem-solving skills while at the same time avoiding details that would favor individuals with prior business experience. The 6 scenarios we developed are briefly summarized in Table 1.

<table>
<thead>
<tr>
<th>Scenario title</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1: personnel shortage</td>
<td>You are the human resource manager of a manufacturing plant facing a personnel shortage. Your employees are working excessive amounts of overtime and morale is low. Materials provided: Current employment figures and job-satisfaction survey results.</td>
</tr>
<tr>
<td>S2: strategic decision making</td>
<td>You are a management consultant working for a newspaper that is struggling to gain new subscribers and is entering the world of electronic media. Materials provided: Memos pertaining to a potential union strike, summary of an evaluation of the web-based paper, and the newspaper’s third quarter financial report.</td>
</tr>
<tr>
<td>S3: problem subordinate</td>
<td>You are the director of research and development for a communications technology company. You have a subordinate manager who is technically sound but lacks managerial skills. Materials provided: Organizational chart, resume and evaluations of the problem subordinate, and memos and email exchanges among project managers and the director.</td>
</tr>
<tr>
<td>S4: consulting challenge</td>
<td>You are a financial consultant to a medium-sized beverage company. The company’s president has sought your services and you discover that he is part of the problem. Materials provided: Press releases, industry performance and market share data, net income chart, and transcripts of interviews with top management.</td>
</tr>
<tr>
<td>S5: inter-departmental negotiations</td>
<td>You are a director of a nonprofit agency and in charge of developing the floor plan for a new building. One of your colleagues is creating conflict by refusing to compromise. Materials provided: Resume of the design consultant, memos regarding floor planning process, email and voice-mail communications among planning team members.</td>
</tr>
<tr>
<td>S6: project management</td>
<td>You are a project manager in a technology-consulting firm. You recently were given additional responsibilities in preparation for a promotion. In addition, your teams are behind on several projects, with critical deadlines approaching. Materials provided: Organization charts, project calendar, communications among project staff and clients.</td>
</tr>
</tbody>
</table>
Respondents were instructed to read through all the materials and to answer a set of questions aimed at assessing the following problem-solving skills:

- **Problem identification and rationale.** What do you see as the main problem in this situation? Why do you consider it to be the main problem? What additional problems need to be addressed?
- **Solution generation and rationale.** What would you do to address the main problem you have identified? What alternative courses of action did you consider? Why did you choose your particular course of action?
- **Information processing.** What information did you focus on in developing a response to the situation? How did you use the information to arrive at a response to the situation? Did you draw on any personal experiences in developing a response to the situation? If so, please explain. What additional information/resources would you need to address this problem?
- **Outcome monitoring and obstacle recognition.** What outcome do you hope will result from the course of action you have chosen? What obstacles, if any, do you anticipate to obtaining this outcome?

Next, 27 business-school alumni and students provided answers to the scenarios and rated them in regard to: (a) time requirements, (b) realism, (c) accuracy and sufficiency of information, (d) prerequisite knowledge or experience, and (e) types of skills/abilities addressed. It took the reviewers, on average, 20–40 min to complete each scenario. Overall, the reviewers believed that the scenarios were realistic, presented clear problems, and provided enough information to generate an appropriate response. Mean ratings were all above 3.5 on a scale of 1 (strongly disagree) to 5 (strongly agree). Next, although reviewers considered the scenarios to be challenging, the majority thought that prior knowledge or experience in a specific management function (e.g., marketing, accounting) was not necessary to respond effectively to the problems. Lastly, reviewers were asked to indicate which of nine leadership skills were addressed by each scenario (see Table 2). As expected, the nine leadership skills were well represented, suggesting that the CSPs provide good coverage of skills that are relevant to effective leadership. When asked to express in their own words what skills and/or abilities were measured by the CSPs, the reviewers mentioned creative thinking, adaptability, common sense, and analytical thinking, among others, thus providing support for the face validity of the items.

The final step of instrument development was to generate rating scales for the open-ended questions of the CSPs. To do so, we compiled all the responses provided by the initial sample of 27 alumni and student

<table>
<thead>
<tr>
<th>Skill/ability</th>
<th>S1%</th>
<th>S2%</th>
<th>S3%</th>
<th>S4%</th>
<th>S5%</th>
<th>S6%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning, organizing, and coordinating</td>
<td>25</td>
<td>18.8</td>
<td>43.8</td>
<td>36.4</td>
<td>90.9</td>
<td>90</td>
</tr>
<tr>
<td>Decision making</td>
<td>100</td>
<td>100</td>
<td>68.8</td>
<td>90.9</td>
<td>81.8</td>
<td>90</td>
</tr>
<tr>
<td>Managing projects/subordinates</td>
<td>25</td>
<td>6.3</td>
<td>100</td>
<td>9.1</td>
<td>63.6</td>
<td>100</td>
</tr>
<tr>
<td>Adapting to changing work demands</td>
<td>62.5</td>
<td>31.3</td>
<td>37.5</td>
<td>9.1</td>
<td>18.2</td>
<td>90</td>
</tr>
<tr>
<td>Collecting and interpreting data</td>
<td>75</td>
<td>100</td>
<td>25</td>
<td>90.9</td>
<td>27.3</td>
<td>0</td>
</tr>
<tr>
<td>Handling stress</td>
<td>25</td>
<td>12.5</td>
<td>43.8</td>
<td>0</td>
<td>45.5</td>
<td>100</td>
</tr>
<tr>
<td>Developing subordinates</td>
<td>6.3</td>
<td>6.3</td>
<td>100</td>
<td>9.1</td>
<td>18.2</td>
<td>40</td>
</tr>
<tr>
<td>Communicating and informing others</td>
<td>37.5</td>
<td>37.5</td>
<td>50</td>
<td>54.5</td>
<td>100</td>
<td>70</td>
</tr>
<tr>
<td>Interacting effectively with others</td>
<td>62.5</td>
<td>31.3</td>
<td>93.8</td>
<td>27.3</td>
<td>100</td>
<td>70</td>
</tr>
</tbody>
</table>

Cell values represent the percentage of reviewers who indicated that the scenario addressed the ability.
reviewers into a survey. A set of expert reviewers, hand-picked for each scenario based on their background and functional area, were asked to evaluate the quality of these responses relative to the problem-solving dimension they represented (i.e., problem identification, solution generation, information processing, and outcome monitoring). Ratings were obtained from 8 to 10 experts per scenario (total \( n = 55 \)). These ratings were used to identify benchmarks of more and less effective answers to include, along with anchors, for 5-point rating scales developed for each problem-solving dimension (e.g., problem identification, solution generation, information use, outcome identification) as well as overall quality.

5.2. Measures of tacit knowledge

Past effort to develop measures of TK have involved soliciting critical incidents from individuals with domain-relevant experience, and developing written descriptions that summarize the main problem in

**Scenario 1: Personnel Shortage**

You are a senior level manager in the human resources department of a medium-sized manufacturing plant (2,500 employees). Your primary responsibility is to oversee employee selection and staffing. The plant has found itself in a unique situation in which product demand has been high but unemployment levels are low. This situation has resulted in a personnel shortage in key areas of the plant (20% in production, 15% in maintenance, and 25% in engineering). To avoid layoffs and reduce overhead costs, the company has previously used temporary laborers to compensate for fluctuations in product demand. For the past six months, product demand has been very high and future projections continue to be positive for the next three to six months. In the short term (three months or less), temporary workers are more cost effective, however, their commitment to the job and work quality is less than full-time employees. In the long-term (six months or more), hiring full-time employees is more cost effective. However, if production demands drop, as they often do, the plant would have to lay off employees, which it has never done in its entire 25-year history.

Given this situation, rate the effectiveness of the options below using the following scale:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extremely Bad</td>
<td>Bad</td>
<td>Somewhat Bad</td>
<td>Neither Good Nor Bad</td>
<td>Somewhat Good</td>
<td>Good</td>
<td>Extremely Good</td>
</tr>
</tbody>
</table>

___ Hire temporary employees to compensate for the immediate shortage and reassess the situation in three months.

___ Hire full-time employees but let them know that if production demands decrease, you will have to let them go.

___ Hire a few full-time employees to fill some of the positions and fill the rest with temporary employees to minimize layoffs should production demand diminish.

___ Ask each department to evaluate their own personnel needs and recommend the best approach for their own department.

___ Research the situation in more detail to get a better indication of future product demand and of the relative costs and benefits of various staffing options before making any final decisions.

___ Present the available information to top management and have them make a final decision on how to best handle the personnel shortage.

___ Offer overtime hours for existing employees to see if they would like the opportunity to make more money before hiring temporary laborers or full-time employees.

Fig. 1. Sample situational-judgment problem (SJP).
those incidents (Hedlund et al., 2003; Wagner, 1987; Wagner & Sternberg, 1985). Because our goal was to measure more general knowledge (or common sense) and to be able to compare the knowledge-based approach to a skill-based approach to measuring practical intelligence, we employed a different method of item development.

First, we identified a set of sub-problems in each of the 6 case scenarios described above. For example, in Scenario 1, we isolated problems pertaining to understaffing, employee turnover, and job satisfaction. Although in the overall scenario these problems were all symptoms of a larger problem (e.g., a mismatch between employee values and management policies), they all can be viewed as significant problems themselves. For each sub-problem, we developed a brief description that attempted to capture, in a paragraph, the key issues pertaining to that problem. We then generated a set of possible solutions for each problem, using the reviewers’ responses to the question “What would you do to address the main problem you have identified?” For example, suggestions for dealing with the problem of understaffing included hiring temporary employees, offering overtime to current employees, asking individual departments to evaluate their personnel needs, and hiring full-time employees with the understanding that they may be laid off if demand decreases. Because this process differed from previous efforts to develop TK tests as well as other situational-judgment inventories, we differentiate our measures by labeling them situational-judgment problems (SJPs). The process of developing these items resulted in a total of 18 SJPs (three problems representing each of the six scenarios). An example of a SJP is shown in Fig. 1.

To develop a scoring key for the SJPs, we compiled all 18 problems into one instrument and asked a sample of 30 experts to rate the options on a scale from 1 (extremely bad) to 7 (extremely good). Again, the experts were selected to represent a variety of backgrounds and leadership positions. The mean ratings of all the experts on each response option served as the scoring key for the SJPs.

6. Study 1

6.1. Methodology

6.1.1. Sample

The initial validation of the PI measures took place with a sample of 422 incoming MBA students at the University of Michigan Business School during orientation in the Fall of 1999. The sample consisted of 313 (74%) males and 109 (26%) females. The racial/ethnic composition was 28 (7%) African American, 124 (29%) Asian, 213 (50%) Caucasian, 22 (5%) Hispanic, and 26 (6%) classified as “Other.” Nine (2%) did not provide information regarding race/ethnicity. The majority (81%) of the students was between the ages of 26 and 35 and none were over age 45.

6.1.2. Measures

Two methods of measuring PI were administered to the students, the CSPs and the SJPs. Due to the time constraints, the number of CSPs administered to each student was limited to two. Each student also was administered six SJPs representing two alternate scenarios. This resulted in three different pairings of CSPs and SJPs (1 and 6; 2 and 5; 3 and 4). We paired scenarios based on their time requirements, level of difficulty, functional areas addressed, and leadership skills measured. We also alternated the order of the CSPs and SJPs so as to control for potential effects of fatigue and time constraints on performance. Each student received one score for the SJPs and another for the CSPs. Scores on the SJPs
represented the similarity of the student’s ratings to those of the experts across all response options. For our purposes, we chose a similarity index (Pearson correlation), which looks at how closely the profile of ratings provided by the student relates to the profile provided by the experts. The Pearson $r$ was preferred as it allowed for more straightforward interpretation of the results and enabled comparisons to be made across SJP.$^1$ Scores on the CSPs reflect the overall quality rating of an individual’s answers to a series of open-ended questions. Ratings were based on a scoring key derived from expert judgments and ranged from 1 (poor answer) to 5 (excellent answer).$^2$

In addition to the PI measures, we administered measures of student demographics (age, gender, race/ethnicity, citizenship), prior work experience, and prior academic performance (undergraduate GPA, GMAT score). Where available, we used GMAT scores and undergraduate GPA from business-school records to avoid potential unreliability in the self-report data. Finally, a reaction questionnaire was administered after each test format, which asked students to evaluate the measures on five dimensions: overall quality, self-assessed performance, job-relatedness, face validity, and potential use.

Several criteria were used to assess student success. Performance in the MBA program was measured using GPA at the end of the 1st year and the end of the program, as well as a score on an applied team-consulting project completed during the student’s first year. The consulting project was evaluated on a scale of 1=Low Pass, 2=Pass, 3=Good, or 4=Excellent. Success outside the classroom was assessed using indicators of participation in extracurricular activities, including participation in student clubs, involvement in volunteer organizations, and leadership positions held while in the MBA program. Finally, we obtained initial indicators of employment success, including the number of interviews and offers received for internships and jobs, as well as starting salary.

6.1.3. Procedure

The PI measures were administered on the first day of orientation week for incoming MBA students. The researchers provided a brief overview of the study’s purpose and then assigned students to one of six classrooms. Second-year MBA students served as proctors and administered the measures to the students. Before completing the assessment, students read and signed a consent form to allow members of the research team to access prior and future records of performance (e.g., GMAT scores, job placement, MBA grades).

6.2. Results

Table 3 presents the descriptive statistics for all study variables. The average score on the SJP was .66, indicating that students’ ratings, on average, correlated fairly well with the experts’ ratings. Because scores on the SJP represented Pearson $r$s, we assessed reliability using Spearman–Brown split-half coefficients on the scores students received on the two sets of SJP they completed. Reliabilities ranged from .61 to .73 across the three versions that were administered. These values are consistent with the reliabilities typically observed for situational-judgment tests, in general, and have been viewed as indicators of the multidimensional nature of such measures (Chan & Schmitt, 1998; Oswald et al., 2004). Scores on the CSP represent the average quality rating an individual received on the two scenarios she or

---

$^1$ The SIP scores were transformed for the purpose of analysis into Fisher $z$ values in order to correct for non-normality in the distribution of correlation scores.

$^2$ Because ratings on all the CSP dimensions were highly intercorrelated, we chose to use the overall quality ratings in our analyses for parsimony.
he completed, and ranged from 1 (poor) to 5 (excellent). The mean score on the CSPs was 3.05, indicating that students, on average, received a satisfactory rating. Intraclass correlation coefficients were used to assess the reliability across raters. The inter-rater reliabilities ranged from .62 to .79, with an average of .70 across CSPs. The average grade on the consulting project was 3.10 on a 4-point scale, with 65% of the students receiving a rating of 3 or “good,” suggesting a possible leniency effect in these ratings. Finally, it is important to note that because many of the extracurricular and placement variables were skewed (e.g., the majority of students had 0 as an answer), these measures were log transformed prior to the analyses.

6.2.1. Predictive validity

We first performed correlational analyses to determine the predictive validity of SJP and CSP scores relative to other predictors and the various performance criteria (see Table 4). Scores on both PI measures were predictive of academic success. Students with higher scores on the SJPs had significantly higher 1st year and final GPAs ($r = .18$ and $.21$, respectively), and also received higher grades on the team-consulting project ($r = .17$). Similarly, students with higher scores on the CSPs had significantly higher 1st year and final GPAs ($r = .21$ and $.30$, respectively) and higher consulting project grades ($r = .17$).

Both GMAT scores and undergraduate GPA also were significant predictors of 1st year GPA ($r = .44$ and $.30$, respectively) and final GPA ($r = .40$ and $.32$, respectively). However, GMAT scores did not correlate significantly with the consulting project grade ($r = .06$, ns). Prior work experience did not correlate with MBA grades, and actually exhibited modest negative correlations with PI measures ($r = -.12$) and undergraduate GPA ($r = -.22$).

SJP and CSP scores exhibited modest, but significant correlations with involvement in extracurricular activity. Students who scored higher on the SJPs participated in more student clubs ($r = .15$) and held more leadership positions ($r = .11$). Students with higher CSP scores held more leadership positions ($r = .18$).

3 All correlations reported in the text are significant at $p < .05$ unless otherwise indicated.
Table 4
Correlations among predictor and outcome variables (Study 1)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Prior work experience</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. GMAT score</td>
<td>.06</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Undergraduate GPA</td>
<td>-22**</td>
<td>.18**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. SJP score</td>
<td>-12*</td>
<td>.08</td>
<td>.03</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. CSP score</td>
<td>-12*</td>
<td>.04</td>
<td>.12*</td>
<td>.29**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. 1st year GPA</td>
<td>-.00</td>
<td>.44**</td>
<td>.30**</td>
<td>.18**</td>
<td>.21**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Consulting project grade</td>
<td>-.05</td>
<td>.06</td>
<td>.15*</td>
<td>.17**</td>
<td>.17*</td>
<td>.12*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Final GPA</td>
<td>.00</td>
<td>.40**</td>
<td>.32**</td>
<td>.21**</td>
<td>.30**</td>
<td>.85**</td>
<td>.36**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Club participation</td>
<td>-.10</td>
<td>-.13*</td>
<td>-.06</td>
<td>.15**</td>
<td>.05</td>
<td>-.08</td>
<td>.02</td>
<td>-.06</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Volunteer participation</td>
<td>-.01</td>
<td>-.12*</td>
<td>-.02</td>
<td>.07</td>
<td>.03</td>
<td>-.11*</td>
<td>-.00</td>
<td>-.08</td>
<td>.12*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Leadership positions</td>
<td>-.13*</td>
<td>-.05</td>
<td>-.01</td>
<td>.11*</td>
<td>.18**</td>
<td>-.01</td>
<td>-.09</td>
<td>.05</td>
<td>.19**</td>
<td>.07</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Internship interviews</td>
<td>-.05</td>
<td>.08</td>
<td>.01</td>
<td>-.03</td>
<td>-.09</td>
<td>-.04</td>
<td>-.10</td>
<td>-.05</td>
<td>.00</td>
<td>-.06</td>
<td>.05</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Internship offers</td>
<td>-.06</td>
<td>-.03</td>
<td>-.02</td>
<td>.01</td>
<td>-.01</td>
<td>-.03</td>
<td>-.05</td>
<td>.07</td>
<td>.02</td>
<td>.03</td>
<td>.11</td>
<td>.54**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Job interviews</td>
<td>.06</td>
<td>-.01</td>
<td>.01</td>
<td>-.04</td>
<td>-.05</td>
<td>-.08</td>
<td>-.14*</td>
<td>-.16*</td>
<td>.07</td>
<td>-.11*</td>
<td>.03</td>
<td>.23**</td>
<td>.04</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Job offers</td>
<td>-.05</td>
<td>.01</td>
<td>.04</td>
<td>.02</td>
<td>.11*</td>
<td>-.03</td>
<td>-.08</td>
<td>-.06</td>
<td>.06</td>
<td>.01</td>
<td>.16**</td>
<td>.11</td>
<td>.22**</td>
<td>.54**</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>16. Base salary</td>
<td>.08</td>
<td>.13*</td>
<td>.07</td>
<td>-.02</td>
<td>.09</td>
<td>.03</td>
<td>-.11</td>
<td>-.01</td>
<td>-.06</td>
<td>.04</td>
<td>.03</td>
<td>.14*</td>
<td>.19**</td>
<td>.03</td>
<td>.06</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Square—root transformations were performed on club participation, volunteer participation, and leadership positions in order to normalize the data prior to analysis.

*p < .05; **p < .01.
GMAT scores correlated with extracurricular activity, but in the opposite direction. Specifically, students with higher GMAT scores participated in fewer student clubs ($r = -0.13$) and volunteered less ($r = -0.12$).

Finally, there were few significant predictors of internship or job placement success as measured by the number of interviews, number of offers, and base salary. Students with higher CSP scores received more full-time job offers ($r = 0.11$) and students with higher GMAT scores received higher base salaries ($r = 0.13$). The latter finding may be attributable to the fact that many employees look at the GMAT scores of job applicants, which ultimately may influence the salary they offer.

### 6.2.2. Incremental validity

Next, we sought to determine the validity of the SJPs and CSPs relative to the two most commonly used criteria in MBA admissions, undergraduate GPA and GMAT scores. First we observed that neither SJP nor CSP scores correlated significantly with GMAT scores ($r = 0.08$ and $0.04$, respectively, ns), suggesting that the PI measures tap distinct abilities from the GMAT. Similarly, SJP scores did not correlate with undergraduate GPA ($r = 0.03$, ns), but CSP scores exhibited a modest correlation with undergraduate GPA ($r = 0.12$).

We conducted a series of hierarchical regression analyses in order to determine the extent to which SJP and CSP scores explained individual differences in MBA performance beyond undergraduate GPA and GMAT scores. In the following hierarchical regressions, we entered GMAT scores and undergraduate GPA in the first step, followed in the second step by either SJP scores, CSP scores, or both PI measures (see Table 5). Based on the zero-order correlations, we chose to focus only on 1st year and final GPA in the hierarchical regression analyses.

Both SJP and CSP scores accounted for significant variance in MBA grades beyond GMAT scores and undergraduate GPA. In predicting 1st year GPA, SJP and CSP scores each accounted for an additional 3% of variance beyond GMAT and undergraduate GPA. In predicting final GPA, SJP and CSP scores accounted for 4% and 6% incremental validity, respectively.4

We also compared CSP and SJP scores directly and found that CSP scores were better predictors of final GPA than were SJP scores ($\beta = 0.21$ and $0.14$, respectively). In addition, CSP scores accounted for an

---

4 The calculation of final GPA includes grade received for the team-consulting project.
additional 4% variance in final GPA after accounting for SJP scores. In comparison, SJP scores only accounted for 2% variance beyond CSP scores. There were virtually no differences between SJP and CSP scores in regard to predicting 1st year GPA. Finally, as shown in Table 5, when both scores were considered together, they explained as much as 8% additional variance in MBA grades beyond GMAT scores and undergraduate GPA.

6.2.3. Group differences

Because there is evidence that females and African Americans score significantly lower on the GMAT than do other gender and racial/ethnic groups (Dobson et al., 1999; Hancock, 1999), we assessed the extent to which group differences emerged in scores on the SJP s and CSP s, and compared them to differences in GMAT scores. First, we found significant gender differences for GMAT scores, SJP scores, and CSP scores. Males scored .34 standard deviations higher than did females on the GMAT (Ms = 681 and 666, respectively, \( t = -3.05, p < .01 \)). Females, however, scored .24 standard deviation higher than did males on the SJP s (Ms = .68 and .65, respectively, \( t = 2.29, p < .05 \)) and .39 standard deviation higher on the CSP s (Ms = 3.26 and 2.98, respectively, \( t = 3.53, p < .01 \)). Using our new measures in addition to the GMAT thus helps “cancel out” the gender differences obtained when only the GMAT is used in assessment.

Second, in regard to race/ethnicity, individuals were classified based on self-report data as African-American, Asian, Caucasian, Hispanic, or Other. The results of ANOVAs to test for group differences are presented in Table 6. We computed \( d \) values for all groups relative to Caucasians. African Americans scored 1.24 standard deviations lower than did Caucasians on the GMAT. In contrast, African-Americans scored only .14 of a standard deviation lower on the SJP s and .42 of a standard deviation lower on the CSP s than did Caucasians.

Other patterns of disparities emerged on the PI measures that were not found on the GMAT. Overall, the racial/ethnic-group differences on the SJP s were modest, with the most notable disparity between Caucasians and Asians (\( d = -.38 \)), favoring Caucasians. Differences on the CSP s were more

<table>
<thead>
<tr>
<th></th>
<th>AfrAm</th>
<th>Asian</th>
<th>Caucasian</th>
<th>Hispanic</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMAT Mean</td>
<td>625</td>
<td>684</td>
<td>681</td>
<td>665</td>
<td>667</td>
</tr>
<tr>
<td>SD</td>
<td>48</td>
<td>41</td>
<td>43</td>
<td>46</td>
<td>36</td>
</tr>
<tr>
<td>( F )</td>
<td>11.58</td>
<td>.07</td>
<td>.07</td>
<td>.36</td>
<td>.31</td>
</tr>
<tr>
<td>( p )</td>
<td>.00</td>
<td>.01</td>
<td>-.02</td>
<td>.09</td>
<td>.05</td>
</tr>
<tr>
<td>( d )</td>
<td>-1.24</td>
<td>.14</td>
<td>-.38</td>
<td>.09</td>
<td>-.15</td>
</tr>
</tbody>
</table>

Table 6
Differences in test scores as a function of self-reported race/ethnicity (Study 1)
pronounced, with the largest disparities between Caucasians and Asians \((d=-.63)\), favoring Caucasians, followed by Caucasians and Hispanics \((d=-.57)\), also favoring Caucasians.

One possible explanation for the pattern of disparities on the SJPs and CSPs is that not all students were native English speakers. In fact, 34% of the students in the sample classified themselves as non-US citizens. These students may have had more difficulty successfully completing the measures, given the extensive reading and writing requirements. After controlling for citizenship status, differences in scores among racial/ethnic groups on the SJPs were reduced to nonsignificance. Differences in CSP scores remained, but were somewhat mitigated. Most notably, the disparity in scores between Asians and Caucasians was reduced from −.63 to −.24. Interestingly, differences in scores among racial/ethnic groups on the GMAT were only exacerbated after controlling for citizenship status. The disparity between African-Americans and Caucasians increased to 1.43 standard deviations, which represents a disparity 3 times greater than that for the CSPs. Overall, these findings indicate that the SJPs and CSPs mitigate some of the disparities inherent in the GMAT, although it is clear that the CSPs present some of their own disparities that need to be addressed in further test development.

6.2.4. Student reactions

The final criterion we examined was how the students perceived the PI measures. Students rated the CSPs and SJPs in terms of overall quality (16 items), self-assessed performance (3 items), job-relatedness (5 items), face validity (4 items), and potential use (4 items). Items were rated on a 5-point scale (1=Strongly Disagree to 5=Strongly Agree). The results of these ratings are summarized in Table 7.

The ratings of the SJPs and CSPs were somewhat mixed. The mean ratings for overall quality were 3.65 for the SJPs and 3.50 for the CSPs. In general, students enjoyed completing the problems, found them to be interesting and challenging, and preferred them to other assessments they have completed. Students also generally felt that they answered the questions effectively and performed as well as or better than did others in their cohort \((M=3.35\) and 3.23 for SJPs and CSPs, respectively). Ratings regarding job-relevance and potential use, however, were more nearly neutral, indicating that there was neither agreement nor disagreement about the relevance of the new measures to MBA admissions or job performance. The lowest ratings were for face validity \((M=2.71\) and 2.73 for SJPs and CSPs, respectively).

<table>
<thead>
<tr>
<th>Format/dimension</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Situational-judgment problems</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall quality*</td>
<td>346</td>
<td>3.65</td>
<td>0.46</td>
<td>.81</td>
</tr>
<tr>
<td>Self-assessed performance*</td>
<td>343</td>
<td>3.35</td>
<td>0.53</td>
<td>.55</td>
</tr>
<tr>
<td>Job relatedness*</td>
<td>343</td>
<td>2.92</td>
<td>0.71</td>
<td>.80</td>
</tr>
<tr>
<td>Face validity</td>
<td>340</td>
<td>2.71</td>
<td>0.71</td>
<td>.81</td>
</tr>
<tr>
<td>Potential use*</td>
<td>340</td>
<td>2.86</td>
<td>0.75</td>
<td>.80</td>
</tr>
</tbody>
</table>

**Case scenario problems**

| Overall quality              | 378 | 3.50 | 0.48 | .80   |
| Self-assessed performance    | 369 | 3.23 | 0.60 | .65   |
| Job relatedness              | 369 | 3.06 | 0.71 | .82   |
| Face validity                | 360 | 2.73 | 0.76 | .83   |
| Potential use                | 355 | 2.95 | 0.75 | .80   |

*Indicates significant mean difference as a function of measurement type.
respectively), indicating that students did not necessarily believe that the measures were fully valid indicators of their own abilities. This finding may reflect the novelty of the measures. That is, students may be reluctant to rate an assessment as a valid measure of their ability without knowing how well they actually scored on that assessment.

We also compared the students’ reactions across the two measures. Significant differences were observed for all dimensions except for face validity. In general, students believed that the quality of the items and their own performance was higher on the SJPs. The CSPs, however, were seen as having greater job-relevance and potential use in admissions and selection decisions.

7. Study 2

Study 2 served as a replication of Study 1 with a new sample of MBA students.

7.1. Method

Study 2 followed the methodology of Study 1 only with slight variations as specified below.

7.1.1. Measures

We administered the PI measures in the same format and collected the same performance criteria as in Study 1 with the exception of placement data. Due to poor economic conditions at the time of data collection (Summer of 2002), many job offers were postponed or withdrawn. Thus, we were unable to include this information in our study.

7.1.2. Sample

The sample for Study 2 consisted of 370 students who entered the MBA program at University of Michigan Business School in Fall of 2000. The sample consisted of 262 (71%) males and 106 (29%) females. The racial/ethnic composition was 31 (9%) African-American, 98 (27%) Asian, 191 (52%) Caucasian, 24 (7%) Hispanic, and 18 (5%) classified as “Other.” Eight (2%) did not provide information regarding race/ethnicity. The majority of the sample (81%) was between the ages of 26 to 35, and no students were over the age of 45.

7.1.3. Procedures

The PI measures were administered following the same procedures used in Study 1. However, there were some differences in regard to the conditions of administration. In Study 1 the PI measures were administered as part of the first day of student orientation. In Study 2 they were administered as part of a workshop held by the Office of Career Development in the evening following the first full day of classes. As a result, the sample size was smaller (377 vs. 422) and the testing conditions were less favorable (evening vs. day) in Study 2 than in Study 1.

7.2. Results

Table 8 presents the descriptive statistics for the Study 2 variables. Study 2 students performed comparably to Study 1 students on most of the predictor and outcome variables. The average split-half
reliability on the SJP was .65 and the average inter-rater reliability on the CSPs was .72. The only
noteworthy difference is that Study 2 students scored slightly lower on the CSPs, on average, than did
Study 1 students (\(M=2.86\) vs. 3.05).

7.2.1. Predictive validity

The patterns of correlations observed in Study 2 were generally comparable to those found in Study 1,
with a few exceptions (see Table 9). SJP scores correlated significantly with consulting project grade
\((r=.14)\) and final GPA \((r=.16)\) but did not correlate with 1st year GPA \((r=.08, \text{ns})\). CSP scores
correlated significantly with 1st year GPA \((r=.14)\) and final GPA \((r=.18)\) but only exhibited a
marginally significant correlation with the consulting project grade \((r=.12, p<.10)\).

Table 9
Correlations among predictor and outcome variables (Study 2)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Prior work experience</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. GMAT score</td>
<td>−.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Undergraduate GPA</td>
<td>−.16**</td>
<td>.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. SJP score</td>
<td>.04</td>
<td>−.06</td>
<td>−.05</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. CSP score</td>
<td>−.02</td>
<td>.03</td>
<td>.05</td>
<td>.20**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. 1st year GPA</td>
<td>−.06</td>
<td>.44**</td>
<td>.23**</td>
<td>.08</td>
<td>.14*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Consulting project grade</td>
<td>−.06</td>
<td>.06</td>
<td>.14*</td>
<td>.12</td>
<td>.19**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Final GPA</td>
<td>−.02</td>
<td>.36**</td>
<td>.22**</td>
<td>.16*</td>
<td>.18**</td>
<td>.89**</td>
<td>.48**</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Club participation</td>
<td>.01</td>
<td>.03</td>
<td>.07</td>
<td>−.05</td>
<td>.14*</td>
<td>−.02</td>
<td>.06</td>
<td>−.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Volunteer participation</td>
<td>−.09</td>
<td>.10</td>
<td>.03</td>
<td>.07</td>
<td>.08</td>
<td>−.01</td>
<td>.00</td>
<td>.02</td>
<td>.13*</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Leadership positions</td>
<td>−.07</td>
<td>−.03</td>
<td>.02</td>
<td>.14*</td>
<td>.12*</td>
<td>−.03</td>
<td>.03</td>
<td>.02</td>
<td>.12*</td>
<td>.10</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Internship interviews</td>
<td>.05</td>
<td>.01</td>
<td>.04</td>
<td>−.06</td>
<td>−.02</td>
<td>−.01</td>
<td>−.14*</td>
<td>−.12</td>
<td>.17*</td>
<td>.04</td>
<td>−.01</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>13. Internship offers</td>
<td>−.01</td>
<td>−.07</td>
<td>.03</td>
<td>.01</td>
<td>.18**</td>
<td>.13*</td>
<td>.04</td>
<td>.10</td>
<td>.08</td>
<td>.07</td>
<td>.14*</td>
<td>.42**</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Square-root transformations were performed on club participation, volunteer participation, and leadership positions in order to
normalize the data prior to analysis.

*p < .05; **p < .01.
Similar to Study 1, GMAT scores and undergraduate GPA correlated significantly with 1st year GPA \((r = .44\) and \(.23,\) respectively) and final GPA \((r = .36\) and \(.22,\) respectively). Once again, neither variable correlated significantly with consulting project grade \((r = .00\) and \(.06,\) respectively, ns).

Also consistent with Study 1, students with higher SJP and CSP scores held more leadership positions \((r = .14\) and \(.12,\) respectively). In addition, students with higher CSP scores participated in more clubs \((r = .14,\) In contrast to Study 1, GMAT scores did not correlate with any of the extracurricular activity variables. As indicated above, we were unable to fully examine job-placement success. However, we did collect data on the number of internship interviews and offers received. We found that CSP scores correlated significantly with the number of internship offers received \((r = .18,\) whereas in Study 1 we observed a relationship between CSP scores and job offers.

### 7.2.2. Incremental validity

Similar to Study 1, scores on the SJP and CSPs exhibited no significant relationship with scores on the GMAT \((r = -.06\) and \(.03,\) respectively, ns) or undergraduate GPA \((r = -.05\) and \(.05,\) respectively, ns). SJP and CSP scores also explained variance in MBA performance beyond GMAT scores and undergraduate GPA, but to a lesser degree than in Study 1 (see Table 10). The most variance explained was for final GPA, with SJP scores accounting for 3% variance and CSP scores explaining 2% variance beyond GMAT scores and undergraduate GPA. When both PI measures were considered in combination, the incremental validity in regard to final GPA increased to 4%.

In Study 1, CSP scores appeared to be slightly better predictors of academic success than were SJP scores. In Study 2, SJP scores accounted for 2% of the variance beyond CSP scores while CSP scores accounted for only 1% of the variance beyond SJP scores, but the differences were not statistically significant.

### 7.2.3. Group differences

Consistent with Study 1, we found significant gender differences for GMAT scores and CSP scores. Males scored .53 standard deviation higher than did females on the GMAT \((Ms = 679\) and \(654,\) respectively, \(t = -4.61, p < .01).\) Females scored .29 standard deviation higher than did males on the

<table>
<thead>
<tr>
<th>Table 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hierarchical regression analyses of MBA grades on predictor variables (Study 2)</td>
</tr>
<tr>
<td>Step</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>1. GMAT score</td>
</tr>
<tr>
<td>Undergraduate GPA</td>
</tr>
<tr>
<td>2. SJP score</td>
</tr>
<tr>
<td>1. GMAT score</td>
</tr>
<tr>
<td>Undergraduate GPA</td>
</tr>
<tr>
<td>2. CSP score</td>
</tr>
<tr>
<td>1. GMAT score</td>
</tr>
<tr>
<td>Undergraduate GPA</td>
</tr>
<tr>
<td>2. CSP score</td>
</tr>
<tr>
<td>SJP score</td>
</tr>
</tbody>
</table>

*p < .05; **p < .01.
CSPs \((M_s=3.01\) and \(2.79,\) respectively, \(t=2.48,\ p<.05)\). In contrast to Study 1, there were no significant gender differences on the SJPs \((M_s=.67\) and .66 for females and males, respectively). Similar patterns of racial/ethnic group differences emerged in Study 2 as were found in Study 1 (see Table 11). African-Americans scored 1.00 standard deviation lower than Caucasians on the GMAT. In comparison, African-Americans scored only .29 of a standard deviation lower than did Caucasians on the SJPs and .50 of a standard deviation lower on the CSPs. Once again, differences on the SJPs and CSPs appear to be attributable to discrepancies between Caucasians and all other groups. However, only one of these differences was larger than half a standard deviation. Asians scored .57 of a standard deviation lower than Caucasians on the SJPs.

As with Study 1, we conducted follow-up analyses to test for the influence of citizenship status on racial/ethnic group differences. Once again, differences in SJP scores were reduced to nonsignificance \((F=.99,\ p=.41)\), with the most notable reduction in the disparity in scores between Asians and Caucasians \((d=-.57\) before and .00 after controlling for citizenship). A similar change was observed on the CSPs, with the difference between Asians and Caucasians reducing from \(d=-.46\) before to \(d=-.08\) after controlling for citizenship. Controlling for citizenship, however, had little effect on other group differences.

### 7.2.4. Student reactions

The pattern of ratings across the five dimensions (overall quality, self-assessed performance, job relevance, face validity, and potential use) was consistent with Study 1, but, overall, the mean ratings tended to be slightly lower in Study 2 (see Table 12). Students tended to rate the measures favorably in regard to quality \((M=3.50\) for SJPs and 3.29 for CSPs), and they were more likely to agree than disagree that they answered the questions effectively \((M=3.35\) for SJPs and 3.13 for CSPs). In Study 2, the students were less likely to agree that the problems were job relevant, face valid, or potentially useful for selection or admissions purposes. Again, the lowest ratings were for face validity \((M= 2.50\) for SJPs and 2.37 for CSPs).
In comparing students’ ratings across measures, we again found that the SJPs were rated higher on overall quality and self-assessed performance than were the CSPs. However, in contrast to Study 1, there were no significant differences between the SJPs and CSPs in regard to job-relatedness or potential use.

8. General Discussion

Our research was prompted by concerns regarding the potential over-reliance on conventional standardized testing in graduate admissions, with specific focus on the GMAT in graduate business school admissions. The concerns with standardized tests like the GMAT are that (a) they measure a limited set of skills relevant to success, accounting for only 20% to 25% of the variance in criterion performance, and (b) they produce disparities in scores among gender and racial/ethnic groups. Our goal was not to replace, but rather to supplement, the GMAT by measuring abilities it does not intend to assess. Specifically, we sought to develop measures of practical intelligence (PI) to complement the analytical emphasis of the GMAT. We administered two types of PI measures (knowledge-based situational-judgment problems [SJPs] and skill-based case scenario problems [CSPs]) to two samples of incoming MBA students and evaluated their predictive validity and potential for adverse impact relative to the GMAT. We summarize and discuss the key findings below.

8.1. Adding value beyond GMAT scores

First and foremost, we examined the extent to which the PI measures predicted success in the MBA program. In general, scores on the PI measures correlated significantly with performance in the program as indicated by graduate GPA and faculty evaluations on a team-consulting project. In addition, CSP and SJP scores correlated moderately with indicators of success outside the classroom, such as the number of leadership positions held and the number of internship and job offers received. As expected, GMAT scores and undergraduate GPA correlated significantly with graduate GPA, but GMAT scores failed to correlate significantly with the consulting project grade. The latter results suggest that some aspects of
graduate school performance are not accounted for by GMAT scores and also are consistent with the finding of Peiperl and Trevelyan (1997) that GMAT scores were not predictive of how students performed on group assignments. It is important to note that GMAT scores did correlate significantly with base salary in Study 1, which does support the findings of Kuncel, Hezlett, and Ones (2004) that analytically oriented tests like the GMAT have predictive validity in regard to work-related criteria.

We next addressed the incremental validity of the PI measures relative to GMAT scores and undergraduate GPA. First, we found no significant correlations between GMAT scores and PI scores, suggesting that the SJPs and CSPs are measuring abilities that are distinct from those measured by the GMAT. Second, both the SJPs and CSPs explained significant variance in grades, particularly final GPA, beyond GMAT scores and undergraduate GPA. Although these increments may be considered modest (ranging from 2% to 6%), Oswald et al. (2004) argued that even an increment of 6% has the potential of adding practical value in college admissions. Further, when both PI measures are considered together, they account for as much as 8% of the variance beyond GMAT scores and undergraduate GPA. Thus a combination of the two measures might provide even greater value to MBA admissions.

In general, the findings of Study 2 are comparable to the findings of Study 1. In both studies, the SJPs and CSPs were predictive of academic performance and accounted for significant variance beyond GMAT scores and undergraduate GPA. However, the effect sizes were slightly smaller in Study 2 compared with Study 1. One possible explanation for the smaller effect sizes in Study 2 was differences in the conditions under which the measures were administered. In Study 2 the PI measures were administered during an optional evening workshop following the first full day of classes, whereas in Study 1 they were administered during the daytime as part of new student orientation. Additionally, in Study 1 the researchers provided a brief overview of the study’s purpose whereas no such overview was provided in Study 2. As a result, the students in Study 2 may have been less motivated and may have been fatigued, which could have negatively impacted their performance. Some direct comparisons of the data from Studies 1 and 2 support this explanation. On the CSPs, which require more effort than the SJPs, students in Study 2 scored significantly lower than students in Study 1 (M = 2.86 and 3.05, respectively, t = 3.71, p < .01). In addition, students in Study 2 rated both PI measures significantly lower in terms of overall quality, job relevance, face validity, and potential use than students in Study 1, which may indicate a generally less positive attitude toward the test.

8.2. Reducing the potential for adverse impact

The second main limitation of the GMAT that we attempted to address is the existence of disparities in scores across gender and racial/ethnic groups. Our findings were consistent with previous research in that females scored lower than males and African-American students scored lower than other groups that were tested (Dobson et al., 1999; Hancock, 1999). The largest disparity was between African-American and Caucasian students, with differences equal to or greater than one standard deviation found in both studies.

We also found significant group differences on the new measures, but the differences tended to be smaller in magnitude and dissimilar in pattern to those observed on the GMAT. Gender differences on the PI measures were comparable in magnitude to the GMAT, but in contrast to the GMAT, females scored higher than males on both the CSPs and SJPs. In regard to self-reported racial/ethnic group, Asian students, rather than African American students, had the lowest scores of all groups on the PI measures. These differences were found to be attributable, in part, to citizenship status. After controlling for citizenship status, differences between Asian and Caucasian students were substantially reduced, if not
altogether eliminated. Differences remained between African-American and Caucasian students, but these disparities were less than half the magnitude of those found for the GMAT. Therefore, PI measures have the potential to correct or to at least partially compensate for some of the disparities in GMAT scores, and thus to reduce the possibility of adverse impact in admissions.

8.3. Knowledge vs. skill-based measures of practical intelligence

We chose to explore two ways of measuring practical intelligence. The SJPs were patterned after a knowledge-based approach used in previous research (Hedlund et al., 2003; Wagner, 1987; Wagner et al., 1999). This approach assesses an individual’s ability to identify more and less appropriate responses to brief problem descriptions, in other words, the individual’s tacit knowledge. The concern with using this type of measure was that it could give those applicants with prior managerial experience an unfair advantage in the admission process. The CSPs represented an alternative approach that involves assessing a specific set of skills associated with solving practical problems. These skills include problem identification, information processing, solution generation, and outcome monitoring. The concern with using this type of assessment is that it can be lengthy to administer and difficult to score.

It is clear from our findings that there are some inherent differences between the two PI measures. The CSPs exhibited slightly better predictive and incremental validities than the SJPs in regard to academic performance. The SJPs, on the other hand, produced less disparity in scores than the CSPs. There also were some notable differences in the ratings of the SJPs and CSPs. The students preferred the format of the SJPs and evaluated their own performance to be higher on the SJPs, but they tended to view the CSPs as more relevant to job performance and more potentially useful in an admission process. Finally, prior work experience was unrelated to scores on both measures, thus eliminating the main concern with using the SJPs with this population. This finding also suggests that the SJPs may be measuring general knowledge (e.g., common sense or good judgment) rather than job-specific knowledge, as has been proposed by past research on tacit knowledge.

Unfortunately, the results do not unilaterally favor one PI measure over the other. Instead, they suggest that both PI measures, through further development, may add value the MBA admissions process. We also found that using both measures, along with GMAT scores and undergraduate GPA, further improves the prediction of graduate school performance beyond using either measure alone. This result indicates that the CSPs and SJPs are tapping into somewhat different aspects of practical intelligence, as intended, and that both aspects are relevant to performance.

8.4. Limitations and directions for future research

There are several limitations of our work that need to be addressed through future research. First, our studies were conducted with just one institution and may not be representative of the results that would be obtained with other MBA programs. Depending on the emphasis placed on the development of practical skills, the SJPs and CSPs may be more or less predictive in other MBA programs. However, the predictive validity of the GMAT in our studies were consistent with those found with samples from other institutions, which suggests that our results may generalize to other schools. Nevertheless, further studies need to be conducted to replicate our findings with other MBA programs.

Second, the validation samples involved students who have already been selected on several criteria, including GMAT scores. Thus, our results indicate predictive validity only within a restricted range of
scores. However, because we do not have population estimates for the PI measures, we did not and could not correct for range restriction in our analyses.\(^5\) It is likely that our findings represent conservative estimates of the actual validity of the measures. Ideally, the measures should be validated with a sample from an extensive and broad MBA applicant pool.

Third, group differences emerged on our measures that were unexpected. In particular, we found differences that appeared to be attributable, in part, to the influence of citizenship status (and the background variables for which it is presumably a proxy variable). That is, non-US citizens received lower scores on the measures than did US citizens. These findings may reflect cultural differences in managerial or leadership styles or first-language differences. The sample used to develop the scoring key for the SJP$s and the raters who scored the CSP$s were predominantly US citizens. If there are cultural variations in the proper way to solve business problems, than it is possible that the responses of non-US citizens would be inconsistent with those viewed as appropriate according to our expert standard, thus resulting in lower scores. In an increasingly global economy, the existence of such differences merits further attention, particularly as they relate to graduate admissions.

Fourth, there is a need to conduct more extensive longitudinal research in order fully to assess the predictive validity of the measures relative to work-related criteria. Our studies tracked students only through the first summer after completing the program. In Study 2, our data collection was negatively affected by poor economic conditions that reduced the number of opportunities for MBA graduates. There were preliminary indicators that CSP and SJP scores may predict success after graduation in terms of job search outcomes, but it will be important to determine the extent to which they predict long-term career success.

Finally, it is important to note that this study represents the initial development and administration of PI measures for MBA admissions. In addition, this study involves the first specific attempt to develop a skill-based measure of practical intelligence. In contrast, the GMAT has been in use since 1954. Clearly, further instrument development is needed to improve the quality of the measures. Post hoc analyses indicated that some versions of the PI measures, which included different sets of CSP$s and SJP$s, were more predictive than others. With further item development, it is plausible, although not certain, that the predictive and incremental validity of the SJP$s and CSP$s would improve. It also is worth exploring the possibility of combining the CSP and SJP in some way that maximizes the predictive and incremental validity offered by adding PI to the admission process.

9. Conclusions

We sought to broaden the range of abilities considered in MBA admission beyond those tapped by the GMAT and undergraduate GPA. Our results indicate that measures of PI predict performance inside and outside the classroom, explain individual differences in MBA grades modestly to moderately beyond GMAT scores, predict performance on a team consulting project better than the GMAT, and generally exhibit less disparity across groups than the GMAT. In the increasingly competitive business environment, it is important to understand what it takes to succeed and to know how to identify and

\(^5\) Corrections for range restriction were computed on zero-order correlations involving GMAT scores. These corrections did not result in any substantial increases or decreases in effect sizes.
develop the talent that leads to success. The traditional admissions process, which relies primarily on GMAT scores and undergraduate grades, does not take into account the full range of abilities that are important for success. Our research showed that PI measures have the potential to improve the prediction of success in business school and beyond. Furthermore, PI measures may ultimately prove beneficial to graduate and undergraduate admissions in general.

Acknowledgements

We would like to thank the numerous individuals who contributed to this project, from item development through scoring: Kevin Plamondon, Andrea Sacerdote, Eric Goodrich, Weihua Niu, Melissa Droller, Evonne Plantinga, Mengdan Chu, Kathryn Rado, Julie Goodrich, Lisa Morgan, Donna Vann, and Robert Silaghi. We also thank former Dean Joseph White of the University of Michigan Business School for his support and input on the project.

References


