

The validity of the Major Field Test in Psychology as a programme assessment tool

Shawn P. Gallagher & Shaun P. Cook

The Major Field Test in Psychology (MFT) is a standardised test designed to assess subject mastery at the conclusion of an undergraduate career. Eighty-one graduating majors completed the MFT and 56 of them also took a multiple-choice exam of questions drawn randomly from an introductory psychology test bank. Like the MFT, the constructed exam was divided into four subscales. A second sample of 29 novice majors also completed the MFT. For the advanced majors, total and subscale scores from the two tests were highly correlated (all $p < 0.01$). Advanced majors scored above national norms and were significantly better than the novices. However, few courses could be linked to performance in the associated MFT subscales but general academic indices like grade point average were highly predictive. The novice MFT scores were similar to the norms for advanced majors. We conclude that the MFT measures broad, basic understanding, not advanced mastery.

Keywords: *Major Field Test; programme assessment; undergraduate psychology programmes.*

LARGE RESEARCH INSTITUTIONS have long been compared internationally. The Shanghai-based Academic Ranking of World Universities and the *Times Higher Education* World Universities Rankings are influential but focus on quantifiable factors like research funding and publications and, for the most part, ignore teaching (Altbach, 2012). This is also true for the rankings published by *US News and World Report*, the most influential domestic scales of US institutions, which only assesses learning indirectly, using graduation rates (*US News and World Report LP*, 2013). However, in the face of the recent global recession, all US institutions, especially publicly funded ones, are under increasing pressure to present objective measures of student learning. For better or worse, this kind of assessment, once the domain of the individual schools, may eventually be used to compile national and international rankings of institutions and their major programmes.

In 1991, the American Psychological Association sponsored the St. Mary's Conference on Enhancing the Quality of Undergraduate Education in Psychology. The

conference attendees discussed psychology education in the US and concluded, among other things, that periodic assessment could not only improve the quality of undergraduate psychology programmes but also demonstrate the value of such programmes to accrediting bodies and benefactors (Halpern et al., 1993; McGovern, 1993). The assessment movement has made significant progress in the past two decades and, at the very least, has given faculty and administrators a set of useful lenses through which to review, evaluate, and improve their undergraduate programmes. Assessment experts have identified a number of domains that warrant periodic review including Student Learning, Curriculum, and Programme Resources (Dunn et al., 2007). Of these, Student Learning is probably the most important to demonstrate but the most difficult to quantify. Learning goals may not only vary across institutions but also within institutions as each psychology department attempts to prepare students for a wide variety of career paths.

One company, Educational Testing Services (ETS), produces a standardised

assessment test, a Major Field Test (MFT), for psychology as well as several other undergraduate subject areas. These tests are 'designed to measure the critical knowledge and understanding obtained by students in a major field of study (ETS, 2013)'. A report published by ETS not only suggests that their products can be critical components of institutional assessment but also calls for their regular administration at all accredited US institutions:

Post-secondary education today is not driven by hard evidence of its effectiveness... A nationwide system of accountability needs to be developed within the context of efforts to monitor and improve higher education... We recommend that... accrediting agencies be charged with integrating a nationwide system of assessing student learning into... ongoing reviews of institutions of higher education (ETS, 2006).

For US institutions, especially those who rely on government support and are annually scrutinised by a tax-paying public, the message from ETS is clear: The MFT is not only a useful tool for measuring 'critical knowledge and understanding' but it is also an objective, quantifiable way to for a reviewing body, governmental or otherwise, to address accountability and, possibly, influence university rankings. The aim of this study was to evaluate the validity of the psychology MFT, a standardised multiple-choice test, as a useful tool for internal or external departmental assessment. Additionally, the strategies that we have employed can be applied to any such test that purports to assess advanced understanding in a discipline.

Our psychology department serves over 500 undergraduate majors and aspires to provide each one with a broad disciplinary foundation. One element of our curricular review process includes the periodic administration of the MFT. More than 7000 students at 200 institutions have taken the 140-question multiple-choice test since 2010

(see the Appendix). ETS identifies participating institutions as those with five or more students taking the test in the 2010–2012 period but there is no specific recruitment protocol. Some institutions administer it regularly to all graduating students (Stoloff & Feeney, 2002) but the cost of \$25 US per exam might lead others, like ours, to administer the test only periodically and to only a sample of students. The MFT test questions cover a wide range of topics in psychology, and test results are reported with a single total score as well as subscores for Learning and Cognition (LRN), Perception, Sensory, Physiology, Comparative, and Ethology (PHYS), Clinical, Abnormal, and Personality (CLIN), and Developmental and Social (DEV). Although the objective results generated by a widely-used test give us a convenient set of numbers to analyse and present, a review of the literature prompted us to take a closer look at the validity of the test (Frazier & Edmonds, 2002; Stoloff & Feeney, 2002).

Thousands of psychology majors from hundreds of institutions have taken the MFT, and many departments have used MFT test results to evaluate their programmes and revise their curricula (Dolinsky & Kelley, 2010; ETS, 2013; Frazier & Edmonds, 2002; Stoloff & Feeney, 2002). According to ETS (2005), the MFT is a tool designed to 'assess mastery of concepts, principles, and knowledge expected of students at the conclusion of an academic major in specific subject areas' (p.1). Presumably, performance on the test will improve as a student progresses through the undergraduate major and acquires a deeper and more comprehensive understanding of psychology. Advanced majors should outperform novices. However, the reports from those using the MFT in programme assessment are, at best, inconclusive about the factors that lead to high test scores. It is not surprising that MFT scores consistently correlate highly and significantly with general indicators of academic proficiency like grade point aver-

ages (GPAs) and SAT¹ (College Board) scores (Dolinsky & Kelley, 2010; Stoloff & Feeney, 2002), but some found that these measures have more predictive value than discipline-specific indicators of acquired knowledge like the number of psychology courses taken (Dolinsky & Kelley, 2010; Stoloff & Feeney, 2002). In fact, success on the MFT has been linked to performance in only a few courses (Stoloff & Feeney, 2002) or, as Frazier and Edmonds (2002) concluded, just one. Although some courses cover subjects named in the MFT subscale titles, Stoloff and Feeney (2002) concluded that students who take courses like Abnormal Psychology, Social Psychology, and Developmental Psychology scored no better on the respective subscales than those who do not.

The published literature challenges the assertion that MFT scores reflect knowledge acquired at 'the conclusion of an academic major' in psychology. Although our informal review of past MFT questions suggested that the psychology test had high face validity, it also revealed that many of the questions dealt with foundational material covered in our introductory psychology course (see the Appendix for sample questions). Results found in the published literature (e.g. Frazier & Edmonds, 2002; Stoloff & Feeney, 2002) also suggest that the MFT may be assessing a student's mastery of basic concepts that are then reinforced in advanced courses. We, therefore, hypothesised that MFT total scores and subscores would correlate with results from an assessment test that we constructed with questions randomly drawn from an introductory psychology textbook's test bank. We also hypothesised that, despite the fact that our advanced majors historically perform at or above the national MFT means, a sample of novice psychology majors would also score close to the normative means.

Method

Programme

Our undergraduate programme requires psychology majors to complete a minimum of 33 credit hours in psychology. Most students are between the ages of 18 and 22 years and nearly half complete the undergraduate programme of 120 credits in four years. Typical courses are worth three credit hours and hold three weekly classroom hours over a 14-week period. Lab courses carry four credits and require additional classroom hours that involve running experiments or learning computer applications. All psychology students are required to take a three-credit course in General Psychology and then two sequential four-credit courses in Statistics and Experimental Design. They must then choose at least one four-credit advanced lab course in Child Development, Cognitive Psychology, Learning and Motivation, or Sensation and Perception. Students select the remainder of courses according to their interests and career objectives.

Percentage grades for each course are converted to grade points that fall on a scale from 0 (65 per cent or lower) to 4.0 (95 per cent or higher). These points are then used to calculate a grade point average for all courses (overall GPA) as well as psychology courses in particular (major GPA). Psychology majors who fail to achieve a major GPA of 2.25 by the end of their second year are typically dismissed from the programme. Psychology majors who fail to achieve an overall GPA of 2.0 are not awarded degrees.

Participants

Eighty-one advanced psychology majors, defined as students who were within one month of completing the required minimum of 33 credits hours in psychology ($M=41.1$ credits, $SD=8.27$), took the MFT as well as other assessments and surveys. A second sample of novice majors ($N=29$),

¹ The SAT is a standardised test used for admissions at most undergraduate institutions in the US. It contains three major sections: Critical Reading, Mathematics, and Writing.

having completed no more than six credit hours of psychology (two courses), also took the MFT. We recruited the students two weeks in advance and told them that they were going to take a test of general subject knowledge. They were not instructed to study for the tests and were given course credit for participation, regardless of performance.

Materials

The assessment tests included the web-based version of the MFT and a 100-question introductory psychology final exam (IPFE) that we created with ExamView® software (Version 5.2.0 FS Creations/now eInstruction, Denton TX) that accompanied an introductory psychology textbook by James Kalat (2005). The questions are published in the book's test bank (Meine & Kalat, 2005). This textbook was currently not in use at the university and was chosen because of the author's expertise in the development of standardised tests including the Graduate Record Examination (GRE) in psychology, also produced by ETS (Kalat & Matlin, 2000). The ExamView® software automatically selected 100 multiple-choice questions from a final exam question pool and compiled them in the form of a single exam. We then used the software to convert the exam to an HTML file that could be loaded onto a university server and administered via campus computers. The authors independently categorised each IPFE question into no more than two of the MFT subgroups, and only questions nominated by both authors were included in a given subscale score. Distributions were as follows; 23 questions were classified under LRN, 18 were classified under PHYS, 21 were classified under CLIN, and 17 were classified under DEV. Four questions were scored under two of the four subgroups, and the remaining 25 questions pertained to other areas, such as ethics and methodology, and were used only in the total score calculation. Surprisingly, the proportions of questions allotted to each subgroup in the IPFE were not significantly different

from those of the MFT ($\chi^2(4)=2.06, p>.05$) which also includes questions (about 20 per cent) that are not specific to any subgroup and used only in the total score calculation (ETS, 2005).

Procedure

We administered tests with campus computers during single, uninterrupted blocks of no more than two hours. Although proctors were available, no student requested assistance after receiving instructions and starting a test. We counterbalanced the order of the MFT and IPFE for the advanced majors at intervals of between two and 14 days. Students took the online MFT according to ETS protocol, and results were automatically submitted to and scored by ETS. We then accessed overall and subscale scores through the ETS website. We could not access answers to specific test questions or determine how many questions had been answered. When students completed the IPFE, the software scored the tests and emailed results to the first author. We entered the responses into a spreadsheet application that calculated subscale scores.

Results

Table 1 summarises mean GPA, SAT and MFT scores for the advanced and novice majors. MFT scores are reported on a scale from 120 to 200; raw scores are not provided. For the advanced students, major GPAs ($M=3.14, SD=0.57$) were similar to their overall GPAs ($M=3.15, SD=0.52$), and the mean overall GPA did not differ significantly from that of the previous year's graduates ($z=1.35, p=0.18$, two-tailed). MFT total and subscale scores were significantly correlated with all GPA and SAT measures (Table 2).

The advanced majors ($N=81$) performed well on the MFT ($M=160.7, SD=15.0$, Range: 126 to 196) and the total and subscale score means were all above the corresponding normative means with the DEV subscale mean being the only one not significantly so (all other $p<0.05$, two-tailed $z, N=7077$). Of the advanced majors who completed the

Table 1: Comparison of GPA, SAT, and MFT total and subscale scores for Advanced Majors (>=33 psychology credits completed), Novice Majors (<=6 psychology credits completed), and the Normative Sample (N=7077).

	Advanced Majors M (SD)	Novice Majors M (SD)	Normative Sample M (SD)	Advanced–Novice	Normative–Novice
GPA Overall	3.15 (0.52)	3.29 (0.59)	-	$t(107)=-1.21$	-
SAT Reading	526.0 (79.5)	555.6 (73.1)	-	$t(78)=-1.41$	-
SAT Math	509.2 (89.0)	535.0 (70.9)	-	$t(78)=-1.13$	-
SAT Writing	522.0 (74.2)	543.9 (81.3)	-	$t(71)=-1.06$	-
MFT Total	160.7 (15.0)	152.3 (12.9)	156.1 (15.5)	$t(108)=2.69^*$, $d=0.26$	$z=1.32$
MFT: LRN	59.1 (15.5)	52.0 (12.8)	55.6 (15.2)	$t(108)=2.22^*$, $d=0.21$	$z=1.28$
MFT: PHYS	62.8 (15.5)	56.2 (11.9)	56.1 (14.7)	$t(108)=2.09^*$, $d=0.20$	$z=-0.04$
MFT: CLIN	60.9 (14.9)	53.3 (14.8)	56.4 (15.4)	$t(108)=2.37^*$, $d=0.23$	$z=1.08$
MFT: DEV	57.9 (14.3)	52.5 (15.3)	55.5 (14.8)	$t(108)=1.71^*$, $d=0.16$	$z=1.09$

MFT=Major Field Test; LRN=learning and cognition; PHYS=perception, sensation, physiology, comparative, and ethology; CLIN=clinical, abnormal, and personality; DEV=developmental and social; Major GPA=grade point average for psychology courses; Overall GPA=grade point average for all courses. * $p<0.05$, one-tailed.

Table 2: Advanced Majors: Pearson correlations of Introductory Psychology Final Exam Scores and Academic Scales with MFT scores.

	MFT: Total	MFT: LRN	MFT: PHYS	MFT: CLIN	MFT: DEV
IPFE: Total	.83**	.75**	.71**	.77**	.64**
IPFE: LRN	.79**	.74**	.66**	.71**	.63**
IPFE: PHYS	.71**	.66**	.68**	.63**	.55**
IPFE: CLIN	.72**	.66**	.61**	.69**	.49**
IPFE: DEV	.68**	.56**	.63**	.62**	.55**
Major GPA	.73**	.70**	.56**	.67**	.61**
Overall GPA	.69**	.65**	.53**	.64**	.54**
Major Credits	.41**	.43**	.42**	.30**	.31**
SAT Reading	.59**	.52**	.44**	.50**	.54**
SAT Math	.47**	.44**	.44**	.30*	.41**
SAT Writing	.58**	.50**	.39**	.51**	.46**

IPFE=Introductory Psychology Final Exam based on questions selected randomly from an introductory text (Kalat, 2005) test bank; MFT=Major Field Test; LRN=learning and cognition; PHYS=perception, sensation, physiology, comparative, and ethology; CLIN=clinical, abnormal, and personality; DEV=developmental and social; Major GPA=grade point average for psychology courses; Overall GPA=grade point average for all courses. Major Credits=Psychology Credits Completed.

* $p < 0.05$, ** $p < 0.01$, correlations for corresponding subscale scores are in bold.

Table 3: Novice Majors: Pearson correlations of Academic Scales with MFT scores.

	MFT: Total	MFT: LRN	MFT: PHYS	MFT: CLIN	MFT: DEV
Overall GPA <i>r</i> (29)	.58**	.59**	.21	.45*	.58**
SAT Reading <i>r</i> (18)	.75**	.70**	.65**	.62**	.71**
SAT Math <i>r</i> (18)	.77**	.70**	.79**	.62**	.69**
SAT Writing <i>r</i> (18)	.63**	.70**	.37	.48*	.55*

MFT=Major Field Test; LRN=learning and cognition; PHYS=perception, sensation, physiology, comparative, and ethology; CLIN=clinical, abnormal, and personality; DEV=developmental and social; Overall GPA=grade point average for all courses.

* $p<0.05$, ** $p<0.01$

MFT, 56 also completed the IPFE. Table 2 shows the Pearson correlations between the elements of the MFT and IPFE. The MFT total scores and subtest scores correlated significantly not only with their IPFE counterparts but also with every other element of the IPFE. The IPFE LRN and PHYS subscale scores correlated best with their MFT counterparts. Major GPA, overall GPA, and major credits completed also correlated significantly with all MFT scores (all $p < 0.05$).

All the novice majors ($N=29$), had completed a course in introductory psychology, and 11 of the 29 had also completed a course in child development. For the novice majors, MFT total and subscale means were all significantly lower than those of the advanced majors ($M=152.3$, $SD=12.9$, Range: 131 to 174), although effect sizes were modest (Cohen's d). Novice means were not significantly lower than the normative means provided by ETS (all $p > .05$, one-tailed z). Five novice majors had MFT total scores that placed them above the 80th percentile for the individual national normative scores (ETS, 2013). MFT total scores and most subscale scores were highly correlated with GPA and SAT scores (Table 2).

We performed a stepwise linear regression analysis to predict MFT total scores from Major GPA, Major Credits, SAT Reading, SAT Math, and SAT Writing scores. Overall GPA was highly collinear with Major GPA ($r=.92$) and, therefore, excluded from the analysis. Each of the predictor variables had a significant ($p < .05$) zero-order correlation with MFT total score (Table 2), but only Major GPA ($\beta=.44$), Major Credits ($\beta=.28$), and SAT Writing ($\beta=.29$) had significant partial effects in the full model (all $p < .01$). The model employing only these three predictors accounted for 62 per cent of the variance in MFT total score, ($F(3,51)=27.96$, $p < .001$).

Like Stoloff and Feeney (2002), we examined the effect of completing specific courses on MFT performance for all courses completed by 10 per cent to 90 per cent of

our advanced majors. This analysis did not allow us to look at the effects of the three required courses taken by all advanced students (General Psychology, Statistics and Experimental Design I & II) and we excluded broad content courses that could be applied to all four subscales (History and Systems, Tests and Measurements). Using t -tests to compare the mean scores of those who did and did not take specific courses, we found no content-specific effects for 17 courses (all $p > 0.10$) which, based on course descriptions, should contribute to specific MFT subscale scores (Table 4). The literature provides little evidence linking specific courses to subscale scores but Dolinsky and Kelley (2010) noted dramatic improvements in PHYS subscale scores after the implementation of a required second-year course in physiological psychology. In our case, those who took physiological psychology ($N=29$), outperformed their peers in all MFT measures (all $p < 0.01$, one-tailed t). Those taking physiological psychology also had higher major GPAs ($t(79)=2.34$, $p=0.01$, one-tailed) and completed more credits ($t(79)=5.76$, $p < 0.01$, one-tailed) than their counterparts. A similar broad, nonspecific effect was observed for History and Systems. Both of these courses tend to attract students with graduate school aspirations.

In a final, exaggerated effort to find a link between course content and an MFT subscale score, we decided to compare two specific groups of advanced majors. Students who take physiological psychology often have plans to pursue graduate training in experimental psychology and we were not surprised by the performance of this highly motivated group. We compared these students to a second group of motivated students with postgraduate plans in clinical, rather than experimental, psychology. In our case, such students can be found by identifying those who have taken both abnormal psychology and personality theory, two topics that fall firmly within CLIN subscale of the MFT. We compared students who took both of these courses but not physiological

Table 4: Courses taken by 10 to 90 per cent of advanced majors for which no content-specific effect on MFT was observed. Each course is listed along with the abbreviation of the subscales for which a specific effect might be anticipated. Required courses taken by all students (General Psychology, Statistics and Experimental Design I & II) and those with broad content (History and Systems of Psychology, Tests and Measurements) were not included.

Abnormal Psychology – CLIN
Behavior Modification – LRN, CLIN
Childhood Disorders – CLIN, DEV
Cognitive Psychology – LRN, PHYS
Counseling Strategies – CLIN
Child Development Lab – DEV
Drug Addiction – CLIN, PHYS
Family Systems – CLIN
Health Psychology – PHYS
Human Adjustment – CLIN
Human Relations – CLIN, DEV
Industrial/Organisational Psychology – DEV
Learning and Motivation – LRN, PHYS
Lifespan and Human Development – DEV
Personality Theory – CLIN
Sensation and Perception – PHYS
Social Psychology – DEV

LRN=learning and cognition; PHYS=perception, sensation, physiology, comparative, and ethology; CLIN=clinical, abnormal, and personality; DEV=developmental and social.

psychology ($N=37$, $M=63.27$, $SD=13.4$) to those who took physiological psychology but not abnormal psychology or personality theory ($N=16$, $M=60.81$, $SD=19.9$). Contrary to our hypothesis, those who took physiological psychology but not the two clinical courses had a higher mean CLIN subscale score ($M=66.7$, $SD=11.01$) than those who took the clinical courses but not physiological psychology ($M=58.5$, $SD=15.3$). These two groups did not differ significantly in mean major GPA ($t(51)=1.16$, $p=.25$, two-tailed), overall GPA ($t(51)=.836$, $p=.41$, two-tailed), or SAT scores (all $p>0.1$, two-tailed t). The physiological psychology group did, however, complete more major credits ($t(51)=5.31$, $p<0.01$). For these matched

samples, we suspect that the seemingly paradoxical result is due to exposure to broad, basic content.

Discussion

Our department has been using the MFT for several years as part of our periodic review process. Although the results are objective and simple to compile and present, published reports compelled us to evaluate the test's validity (Frazier & Edmonds, 2002; Stoloff & Feeney, 2002). Previous studies have challenged the notion that multiple choice exams can measure critical thinking skills (Stanger-Hall, 2012) and found that they typically over-estimate general content mastery (Funk & Dickson, 2011). The results

of our investigation are consistent with these findings and have led us to four conclusions. First, MFT total and subscale scores can be predicted by performance on a comprehensive introductory psychology exam. Second, although our advanced students performed well on the MFT, as predicted by Stoloff and Feeney (2002), we found no link between specific courses and MFT subscale scores. Third, novice majors can perform well on the MFT despite having completed no more than two psychology courses. Finally, performance on the MFT appears to be a function of overall academic potential as measured by standardised tests and GPA.

In our opinion, the sample questions provided by ETS (Appendix A) and the questions that we have seen on the actual tests assess basic, introductory-level knowledge. Although we observed significantly different MFT scores between novice and advanced students, the question of what led to these differences remains. We suspect that these differences are due to the fact that the advanced students have had more opportunities to review and discuss basic material common in introductory courses. The students who do well on the MFT might well have a deep understanding of psychology, but they also have had many opportunities to review core content; basic Freudian principles get reinforced through courses in Personality Theory or Abnormal Psychology and basic Neuroanatomy is reviewed in courses like Sensation and Perception. Our top-scoring novices are exceptional students who have performed well in all of their college courses and have probably mastered the material they have encountered so far. However, regardless of how exceptional a novice student happens to be, a test 'designed to assess mastery of concepts, principles, and knowledge expected of students at the conclusion of an academic major' (ETS, 2005) should contain material well beyond his or her reach. Although one of our five top-scoring novice students took an advanced placement psychology course in high school, none of the others had any

supplemental psychology education or training; two had completed only General Psychology, and three had completed General Psychology and Child Development. We suspect that exceptional novice students from any university would do well on the MFT, and the gaps we observed between advanced and novice scores might be even smaller for departments with two-semester introductory psychology requirements. Although they will not have data that they can directly compare to a normative sample, we recommend that, if a department aspires to measure the broad, basic content knowledge of its advanced majors, they construct an IPFE of their own with questions from a comprehensive introductory psychology textbook.

Although our specific claim that the MFT fails to validly assess mastery might be new, our results are consistent with many previous reports. Like Stoloff and Feeney (2002), we found that although psychology credits completed was significantly correlated with MFT total, GPA, either Overall or Major, accounted for approximately 50 per cent of the variance in MFT total while credits completed accounted for less than 20 per cent of the variance. In another study exploring MFT correlates, Frazier and Edmonds (2002) found that, of all the academic indicators evaluated, MFT scores were predicted only by performance in a single in-depth, and closely mentored, research methods course. The authors noted that those who excel in this class are typically those planning on graduate school and, we suggest, that they were the ones who made a point of mastering the basics from the beginning.

In a four-year longitudinal study, Dolinsky and Kelly (2010) noted impressive gains in MFT scores after converting Abnormal Psychology, Physiological Psychology, and History and Systems from optional to required courses. Although it is likely that the course in Abnormal Psychology contributed to the gains in the CLIN subscale scores and that Physiological

Psychology helped the scores on the PHYS subscale, the LRN and DEV subscales exhibited similar gains without any curricular modifications in the associated sub-disciplines. It is possible that the History and Systems course, due to the topic's broad nature, facilitated MFT gains across the four subscales. Although History and Systems should introduce even the most advanced students to new material, it is impossible to teach the course without reviewing key elements from across psychology. Dolinsky and Kelly seem aware of the benefits of reviewing basic material and note that their department briefly contemplated adding a capstone review-and-synthesis course or a 'mini-review' aimed at improving MFT scores. Although they concluded that dedicating a single course to a standardised test would be an inappropriate allocation of resources, it probably would have worked.

Our homemade assessment test, the IPFE, was generated with software that randomly drew questions from an introductory psychology textbook's final exam test bank. Having little knowledge about how MFT questions were selected, we concluded that the ExamView® software would be the least biased arbiter of content. As it turned out, the proportional distribution of content was remarkably close to that of the MFT, possibly due to the fact that the textbook author has been affiliated with ETS. It is, however, noteworthy that the IPFE was much shorter than the MFT and that the strength of the correlation between the subscale pairs was proportional to the number of IPFE questions in the category. We did not want to bias the IPFE by adding or subtracting questions, but we believe that the MFT subscale correlations could have been even higher had we made the IPFE longer and actively managed content with input from across our department.

If the MFT is primarily measuring basic knowledge in psychology, this fact will remain hidden each time a department's top

students live up to faculty expectations by outperforming their peers. To the casual observer, the test appears to be doing what it has been designed to do. However, we have found that MFT scores can be accurately predicted by performance on an introductory psychology exam, the subscale scores cannot be clearly linked to courses, and the test cannot clearly differentiate between novices and a large normative sample of graduating majors. Our department is now faced with the challenge of deciding how these and future MFT results will be used. We will probably continue to employ the test in a limited fashion for the sake of comparative analysis but, in its current state, the MFT will have little impact on our department's curricular development. As a result of this investigation, we have redoubled our efforts to evaluate our own course offerings and requirements. We are also constructing our own assessment tests which will incorporate some multiple choice questions but also include data analysis and critical reviews of published studies. Our test will, of course, be administered at a fraction of the MFT's cost.

Comprehensive and effective programme review is a multi-dimensional endeavor and each component of the process must be closely scrutinised and validated (Dunn et al., 2007). Like many universities, our institution is facing increased pressure, from both internal and external sources, to objectively demonstrate programme effectiveness, and the results from assessment initiatives can have profound implications. Courses can be offered or withdrawn and faculty bodies can expand or contract based on data generated by quantitative instruments like the MFT. Furthermore, the administration of these tests can command a significant amount of money, faculty and student hours, and facility space. Our hope is that the resources dedicated to any assessment initiative are in proportion to its demonstrated validity.

Author Notes

We thank Rich Carlson and Bea Grosh for reviewing and commenting on earlier drafts of this manuscript.

This work was supported in part by the Millersville University Department of Planning, Assessment and Analysis, Dr Lisa R. Shibley, Assistant Vice President.

This work was presented in part at the annual meeting of the Eastern Psychological Association, Cambridge, Massachusetts, 12 March 2011.

References

- Altbach, P.G. (2012). The globalisation of college and university rankings. *Change: The Magazine of Higher Learning*, 44, 26–31.
- Dolinsky, B. & Kelley, J.M. (2010). For better or worse: Using an objective programme assessment measure to enhance an undergraduate psychology programme. *Teaching of Psychology*, 37, 252–256.
- Dunn, D.S., McCarthy, M.A., Baker, S., Halonen, J.S. & Hill, G.W., IV. (2007). Quality benchmarks in undergraduate psychology programmes. *American Psychologist*, 62(7), 650–670.
- Educational Testing Service (2003). Retrieved 20 April 2013, from:
<http://www.ets.org/Media/Tests/MFT/pdf/psyques.pdf>
- Educational Testing Service (2005). Retrieved 20 April 2013, from:
http://www.ets.org/Media/Tests/MFT/pdf/mft_testdesc_psych_4bmf.pdf
- Educational Testing Service (2006). Retrieved 20 April 2013, from:
http://www.ets.org/Media/Resources_For/Policy_Makers/pdf/cultureofevidence.pdf
- Educational Testing Service (2013). Retrieved 20 April 2013, from:
<http://www.ets.org/mft/about>
- Frazier, T.W. & Edmonds, C.L. (2002). Curriculum predictors of performance on the Major Field test in Psychology II. *Journal of Instructional Psychology*, 29(1), 29–32.
- Funk, S.C. & Dickson, L. (2011). Multiple-choice and short-answer exam performance in a college classroom. *Teaching of Psychology*, 38(4), 273–277.
- Halpern, D.F., Appleby, D.C., Beers, S.E., Cowan, C.L., Furedy, J.J., Halonen, J.S., Horton, C.P., Peden, B.F. & Pittenger, D.J. (1993). Targeting outcomes: Covering your assessment concerns and needs. In T.V. McGovern (Ed.), *Handbook for enhancing undergraduate education in psychology* (pp.23–46). Washington, DC: American Psychological Association.
- Kalat, J.W. (2005). *Introduction to psychology* (7th ed.). Belmont, CA: Thomson/Wadsworth.
- Kalat, J.W. & Matlin, M.W. (2000). The GRE psychology test: A useful but poorly understood test. *Teaching of Psychology*, 27, 24–27.
- McGovern, T.V. (Ed.) (1993). *Handbook for enhancing undergraduate education in psychology*. Washington, DC: American Psychological Association.
- Meine, P. & Kalat, J.W. (2005) *Test bank for introduction to psychology* (7th ed.). Belmont, CA: Thompson/Wadsworth.
- Stanger-Hall, K.F. (2012). Multiple-choice exams: An obstacle for critical thinking in introductory science classes. *CBE-Life Sciences Education*, 11, 294–306.
- Stoloff, M.L. & Feeney, K.J. (2002). The Major Field Test as an assessment tool for an undergraduate psychology programme. *Teaching of Psychology*, 29(2), 92–98.
- US News and World Report LP (2013). Retrieved 19 August 2013, from:
<http://colleges.usnews.rankingsandreviews.com/best-colleges>

About the Authors

Shawn P. Gallagher & Shaun P. Cook
Millersville University of Pennsylvania.

Correspondence

Shawn P. Gallagher

Department of Psychology,
145 Byerly Hall,
Millersville University,
P.O. Box 1002,
Millersville, PA 17551.
Email: sgallagher@millersville.edu

Appendix: Major Field test in Psychology sample questions (ETS, 2003).

1. Breland and Breland trained pigs to carry wooden coins in their mouths to a piggy bank. This sequence was reinforced with food. After some weeks, however, the trained pigs began to root the coins with their noses, treating them like pieces of food. This can be best characterised as an example of:
 - (A) avoidance responding;
 - (B) conditional responding;
 - (C) superstitious behaviour;
 - (D) instinctive drift;
 - (E) delayed conditioning.

2. Which of the following therapeutic interventions places the most emphasis on gaining insight into early childhood relationships?
 - (A) Systematic desensitisation.
 - (B) Behaviour modification.
 - (C) Family therapy.
 - (D) Gestalt therapy.
 - (E) Psychoanalysis.

Questions 3 and 4 are based on the following passage.

A psychologist investigated the developmental relationship between the average daily amount of television viewing and the reading skills of children. Parents of children in four age groups (6-year-olds, 7-year-olds, 8-year-olds, and 9-year-olds) were asked to record the number of hours their children watched television for a six-month period. The psychologist also gave the children reading-speed and reading-comprehension tests on a monthly basis for the six-month period. Analyses of the data reveal the following correlations:

Age	Correlation between Hours of Television Viewing and Reading Speed	Correlation between Hours of Television Viewing and Reading Comprehension
6	0.13	-0.32
7	-0.03	-0.38
8	0.07	-0.41
9	-0.05	-0.49

3. The pattern of results above suggests which of the following about television watching?
 - (A) It increases reading comprehension but does not increase reading speed.
 - (B) It has a stronger relationship to reading speed than to reading comprehension.
 - (C) It has a stronger relationship to reading comprehension than to reading speed.
 - (D) It diminishes the relationship between reading speed and reading comprehension.
 - (E) It has a diminishing relationship to reading skills as the child grows older.

4. Based on the correlational data, the psychologist claims that television viewing significantly reduces reading skills. This claim can be justly criticised because:
 - (A) children younger than 10 years prefer television viewing to reading and the sample in the study is, therefore, biased;
 - (B) a cross-sectional study cannot provide information about longitudinal development effects;
 - (C) reading comprehension is more difficult to assess than is the amount of television viewing;
 - (D) television viewing need not impede the acquisition and utilisation of reading skills;
 - (E) correlational data do not justify inferences about causes.

5. A juror in a criminal case believes that the defendant's illegal act cannot be explained or excused by extenuating circumstances. According to attribution theory, the juror is most probably:
 - (A) making stereotypical judgments;
 - (B) emphasising dispositional factors;
 - (C) discounting altruistic motivation;
 - (D) overestimating situational causes;
 - (E) relying on compliance to the law.

6. Lesions in Broca's area of the association cortex are most likely to result in which of the following disorders?
 - (A) Expressive aphasia.
 - (B) Visual agnosia.
 - (C) Apraxia.
 - (D) Agraphia.
 - (E) Alexia.

7. A clinical psychologist is conducting a diagnostic interview with a client. Her impression that the client is suffering from paranoid schizophrenia would be supported by the presence of each of the following symptoms EXCEPT:
 - (A) poor contact with reality;
 - (B) systematised delusions;
 - (C) social withdrawal;
 - (D) panic attacks;
 - (E) loose associations.

8. In the course of learning their first language, young children may say 'goed' (for 'went') and 'man's' (for 'men'). These kinds of errors suggest that young children tend to:
 - (A) pay little attention to what they hear;
 - (B) overgeneralise the regularities that they hear in language;
 - (C) produce words that they do not comprehend;
 - (D) use correct grammatical forms only after they have developed a large vocabulary;
 - (E) imitate the grammatical errors of adults.