General Outcome Measurement of Broad Psychology Learning

FTCP Capstone Project Summary

Kevin J. Filter, Ph.D.
Spring 2005

Introduction

I am developing and evaluating new methods of assessing student learning of psychology that are based on the General Outcome Measurement concept exemplified in curriculum-based measurement (CBM; Shinn, 1989). I would like to establish expected growth rates on these measures and use these expected rates as a basis for formative and summative evaluation of student learning. I consider evaluation based on growth rate to be an important addition to the overall process of summative academic evaluation because it is less sensitive to high baseline levels and provides acknowledgement for the gains of students with low baseline knowledge.

Methods

Students enrolled in my introduction to psychology course were administered two types of exams throughout the term on a regular schedule. Participation was voluntary. The exams used as part of this project were not used as a basis for grades this term as their validity, reliability, and utility have not yet been established. Each type of exam was administered at the beginning of the term and on a monthly basis thereafter. Participants' performance on these exams will be correlated with their grades on regular course exams and their overall grades in the course to determine as one part of determining the validity of these new methods of assessment.

The first type of exam was a 25-point multiple choice test that included randomly selected questions from throughout the curriculum. The questions were drawn from the text publisher's question bank. The exam should take approximately 20 minutes to complete for a total student time investment throughout the term of about 1.5 hours. An example of one of the 25-question exams is in the Appendix.

The second type of exam was a 2-minute timed word recall test. Students were directed to list all of the terms that they can think of related to psychology on a lined sheet of paper (see Appendix).

Results and Discussion

Advanced statistical analysis will occur after all data is collected. Present data, however, indicated a significant positive correlation ($r = .404$) between students average class exams scores and the difference (or "improvement") between their first two GOM multiple choice quizzes (quiz 2 - quiz 1) (Table 1). This indicates that there is a relationship between how well people score on unit exams and the amount of overall information that they have learned. Word list scores, however, did not correlate with either of these variables, suggesting that it is not measuring the same learning that is measured by multiple choice quizzes. Table 2 indicates the correlation between the four

experimental test administrations. The only significant correlation was between the two administrations of the word list (r=.395).

There does not appear to be any correlation between scores on wordlists and scores on GOM multiple choice quizzes. However, each measure was sensitive to an improvement over time. The mean difference between administration #2 and administration #1 for the GOM multiple choice quiz was 3.36 (SD = 3.96). The mean difference between administration #2 and administration #1 for the word list was 5.56 (SD = 5.52).

Table 1: Correlations

		Difference1and2Quiz	Difference1and2Wordlist	AverageExamScore
Difference1and2Quiz	Pearson Correlation	1	-.129	.404(**)
	Sig. (2-tailed)	.	.350	.002
	N	56	55	56
Difference1and2Wordlist	Pearson Correlation	-.129	1	.030
	Sig. (2-tailed)	.350	.	.826
	N	55	55	55
AverageExamScore	Pearson Correlation	.404(**)	.030	1
	Sig. (2-tailed)	.002	.826	.
	N	56	55	96

** Correlation is significant at the 0.01 level (2-tailed).

Table 2: Correlations

		Quiz1	Wordlist1	Quiz2	Wordlist2
Quiz1	Pearson Correlation	1	-.091	.214	-.001
	Sig. (2-tailed)	.	.387	.113	.997
	N	92	92	56	55
Wordlist1	Pearson Correlation	-.091	1	.103	.395(**)
	Sig. (2-tailed)	.387	.	.449	.003
	N	92	92	56	55
Quiz2	Pearson Correlation	.214	.103	1	.014
	Sig. (2-tailed)	.113	.449	.	.913
	N	56	56	61	60
Wordlist2	Pearson Correlation	-.001	.395(**)	.014	1
	Sig. (2-tailed)	.997	.003	.913	.
	N	55	55	60	60

** Correlation is significant at the 0.01 level (2-tailed).

In summary, initial results indicate that each measure appears to capture different aspects of learning and the improvement on GOM multiple choice quizzes correlates

highly with performance on regular class quizzes. Further analyses will need to be conducted when all data are collected to determine the psychometric and conceptual contribution that each could make to classroom assessment.

APPENDIX

4-digit ID # _____

Term Listing
Administration # ___

Directions: When the instructor says "begin," please list all of the terms that you can think of related to psychology in the spaces below. Please use the back side if you need more space. When the instructor says "stop," please put your pen or pencil down and stop writing.

_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____
_____	_____