Comparing Multiple-Choice, Essay and Over-Claiming Formats as Efficient Measures of Knowledge

Kevin M. Williams, Bryce Westlake, & Delroy L. Paulhus

The University of British Columbia

Abstract

Accuracy on over-claiming tests (the ability to distinguish real items from foils) has been shown to be a sensitive measure of knowledge in a number of domains. Here we compared the efficiency of the over-claiming technique relative to multiple-choice and essay/short answer techniques. One-hundred and eight first-year undergraduate students completed a psychology knowledge test including items of all three formats. The over-claiming technique exhibited the highest reliability per unit of administration time. The over-claiming technique was also at least as valid as the other two methods, as illustrated by correlations with university grades, verbal ability, and concurrent validities. Results represent further evidence that the over-claiming technique may be an efficient alternative to traditional methods of knowledge assessment including classroom testing.

Introduction

In the over-claiming (OC) technique of knowledge assessment (Paulhus & Harms, 2004), respondents are asked to rate their degree of familiarity with relevant content items (e.g., names, places, events). Because some of the items (foils) do not actually exist in reality, knowledge accuracy scores can be calculated. Previous studies have demonstrated, when scored on academic content, OC accuracy scores correlate strongly with measures of cognitive ability (e.g., Paulhus & Harms, 2004).

The OC technique shares many of the desirable features of the MC technique (e.g., objective item scoring). Previously, we showed that OC items could be administered at a rate of speed six times that of MC items but showed lower reliability and validity per item. To ensure a more fair comparison in the current study, we compared the methods with respect to reliability and validity per unit of administration time. We also aimed to expand upon previous research findings by comparing Short Answer (SA) items with OC and MC items.

Method

Participants. Participants were 108 undergraduate students enrolled in a first-year psychology course (59% female; 53% East Asian). Each student received course credit for participation.

Materials and procedure: The psychology knowledge test consisted of three counterbalanced sections: OC, MC, and SA. Students completed the test in a supervised laboratory session at their own pace and were timed. Item content covered knowledge of social, personality, developmental, and clinical psychology. All items were developed by the

researchers with the assistance of various expert sources (e.g., psychology study guides, textbooks).

The OC section consisted of 64 real items and 16 foils. The foils were developed by the researchers to appear plausible. Participants rated their familiarity with each item on a 1 (never heard of it) to 5 (extremely familiar) point scale. The 20 MC items included five options. For the twelve SA items, students were asked to write at least three sentences in their answer, and were given a maximum mark of three points per item.

Verbal ability. During the lab session, participants also completed the 50-item UBC Word test (Paulhus, 2003). Items included a stem and four options: The participant is asked to select the most appropriate synonym.

Predictive validity. We used the participants' final grade in their psychology course as a criterion for psychology knowledge. These grades consisted of an aggregate of several exams, which were primarily of MC format but also included some SA items. They contained no OC items.

Results

Note from Table 1 that the reliability was highest for the Short Answer format but so was the administration time. The three reliability values were then projected for 10 min of testing time. This procedure yielded reliability *efficiency* estimates of .78, .63, and .52 for the OC, MC, and SA methods, respectively.

Convergent validity. Correlations among the three knowledge test sections were strong: OC accuracy correlated with the MC and SA scores at .55 and .57, respectively, and MC

correlated with SA at .63 (all $p < .01$). None of these correlations differed significantly from each other.

Verbal ability. The OC, MC, and SA sections correlated .49, .47, and .53 (all $p < .01$) with verbal ability, respectively (Table 2). None of these correlations differed significantly from each other. After equating these correlations per unit time, the validity *efficiencies* were .72, .62, and .53, respectively. These values were each significantly different from each other at $p < .05$.

Predictive validity. Note from Table 2 that the correlations with Verbal Ability were comparable despite the administration time difference. The efficiency values were highest for OC. The formats correlated .37, .42, and .39 (all $p < .01$) with final grades, respectively. None of these correlations differed significantly from each other. Controlling for verbal ability had little effect on these values. After equating the methods for administration time, the validity efficiency estimates were .61, .57, and .20, respectively. OC and MC efficiencies did not significantly differ from each other but were both significantly higher than SA.

Discussion

Our previous research (Williams et al., 2004), showed that the OC technique performed as well or better than the Multiple Choice technique on several important psychometric criteria including reliability, and validity measured with respect to overall course grades. The present study extended that work by showing the superiority of OC to Short Answer format.

The comparison of item format was facilitated by equating the three formats with respect to administration time. Then they could be compared directly with respect to reliability efficiency and validity efficiency. On both criteria, the good performance of the OC technique

became more clear. The format appears to perform at least as efficiently as MC and much more efficiently than Short Answer format. After some further refinements, we are optimistic that the OC technique can be applied to classroom and other large-scale assessment settings.

References

Paulhus, D.L. (2003). *The UBC Word Test.* Unpublished instrument, University of British Columbia.

Paulhus, D.L., & Harms, P.D. (2004). Measuring cognitive ability with the over-claiming technique. *Intelligence, 32,* 297-314.

Williams, K.M., et al. (2004, February). *Examining the Over-Claiming Technique as an alternative to traditional classroom testing.* Paper presented at the 3rd Conry Conference on Measurement, Evaluation, and Research Methodology, Vancouver, Canada.

Table 1. Average administration times and reliability efficiencies.

Item Format	Administration Time (mins)	Alpha Reliability	Reliability Efficiency
Over-claiming	3.8	.48	.78
Multiple-Choice	5.4	.54	.63
Short Answer	24.2	.74	.52

Table 2. Correlations of knowledge tests with verbal ability and course grades.

Item Format	Verbal Ability	Predictive Validity
OC	.49 (.72)	.37 (.61)
MC	.47 (.62)	.42 (.57)
SA	.53 (.53)	.39 (.20)

Note: $N = 108$. Correlations adjusted for administration time appear in parentheses. Values of .26 significant at $p < .01$, 2-tailed.