ED 382 631 TM 023 025

AUTHOR	Harmon, Michelle G.; And Others
TITLE	Confirmatory Factor Analysis of the Gibb Experimental Test of Testwiseness.
PUB DATE	Nov 94
NOTE	27p.; Paper presented at the Annual Meeting of the Mid-South Educational Research Association (Nashville, TN, November 9-11, 1994).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Factor Structure; Higher Education; Models; Sampling; Scores; *Skills; Test Coaching; *Test Wiseness; *Undergraduate Students
IDENTIFIERS	*Confirmatory Factor Analysis; *Gibb Experimental Test of Testwiseness; LISREL Computer Program

ABSTRACT

The stability of a two-factor model recently proposed for the Gibb Experimental Test of Testwiseness was assessed, using confirmatory factor analysis. Designed to measure seven specific testwiseness skills with 10 items per skill, Gibb's test has been shown to discriminate between persons trained and untrained in selected testwiseness skills. Such a measure would have greater utility if the structure of the test were identified. Participants were 173 undergraduates. Confirmatory factor analyses were performed with LISREL 8 using total scores on the seven skills. Results indicated that the data fit the two-factor model and the simpler one-factor model. For this sample, the Gibb test could be characterized as tapping a general proficiency in testwiseness. Confirmation of the parsimonious one-factor model supports use of total scores from Gibb's test, although sampling fluctuation may be a concern. Gibb's test appears amenable to yielding a shorter form with fewer subscores or scales, which should facilitate measurement of testwiseness in future studies of training programs. Three tables present analysis results. (Contains 25 references.) (SLD)

Confirmatory Factor Analysis of the Gibb Experimental Test of Testwiseness

Michelle G. Harmon, David T. Morse, and Linda W. Morse
Mississippi State University

Running head: TESTWISENESS

Abstract

The purpose of the study was to assess the stability of a two-factor model recently proposed for the Gibb Experimental Test of Testwiseness, using confirmatory factor analysis. Designed to measure seven specific testwiseness skills with 10 items per skill, Gibb's test has been shown to discriminate between persons trained and untrained in selected testwiseness skills. Such a measure would have greater utility if the structure of the test were identified.

Participants were 173 undergraduate volunteers who took the Gibb test. Confirmatory factor analyses using LISREL 8 were performed using total scores on the seven skills as data. One- and two-factor models were compared.

Results indicated that the data fit the two-factor model, and the simpler one-factor model. For this sample, the Gibb test could be characterized as tapping a general proficiency in testwiseness.

Implications are: (a) confirmation of the parsimonious one-factor model supports use of total scores from Gibb's test; (b) since the original study did not support a one-factor model and used a comparable population, sampling fluctuation may be a concern; and (c) Gibb's test appears amenable to yielding a shorter form having fewer subscores or scales, which should facilitate measurement of testwiseness in future studies or training programs.

Confirmatory Factor Analysis of the Gibb Experimental Test of Testwiseness

Testwiseness is a construct known to affect the validity of test scores because test-taking skills contaminate and confound the assessment of evaluating acquired knowledge (Thorndike, 1951; Fagley, 1987; Rogers & Bateson, 1991). Millman, Bishop and Ebel (1965, p. 707) defined testwiseness as "a subject's capacity to utilize the characteristics and formats of the test and or the test taking situation to receive a high score. Test-wiseness is logically independent of the examinee's knowledge of the subject matter for which the items are supposedly measures." Dolly & Williams (1986) quoted Ebel (1965) who argued that testwiseness exists differentially for individuals, and that students low in testwiseness skills are at a disadvantage in the testing situation. Others (Masters, 1988; Rogers & Bateson, 1991) have found empirical evidence to support Ebel's statement. With the current interest in competency testing emphasized in government programs such as Education 2000, it becomes increasingly more "important to minimize the measurement errors caused by individual differences in test-taking skills" (Samson, 1985, p. 261). Students low in testwiseness skill need to be identified so that procedures can be taken which will reduce any unfair testing advantage due to testwiseness, and thereby increase the validity of test scores for all students concerned.

What Is Testwiseness and Can It Be Learned?

Millman et al. (1965) presented an analysis of the components of testwiseness. The researchers' intention was to provide a theoretical framework on which to base future studies relative to the significance and role of testwiseness as a strategy. The paper has been described as the "classic theoretical work" (Sarnacki, 1979) guiding testwiseness research. Millman et al. divided the components of testwiseness into two main categories, those independent of the test constructor or test purpose, and those dependent on the test constructor or test purpose. Components independent of test constructor or test purpose included stategies for: (a) using test time wisely, (b) avoiding careless errors, (c) making a best guess, and (d) choosing an answer using deductive reasoning. Components dependent of test constructor or test purpose included strategies for: (a) interpreting the test constructor's intent, and (b) using cues contained within the test itself. Skills which assist the examinee in avoiding the loss of points from variables other than knowledge include: time-using strategies, error-avoidance strategies, and knowing the intent of the examiner. Skills related to gaining points from variables other than knowledge include: guessing, deductive reasoning, and cue-using strategies. All of the components outlined by Millman et al. have served as a framework by which researchers (e.g., Slakter, Koehler, & Hampton, 1970; Diamond & Evans, 1972) have formulated questions to examine testwiseness skill in students.

Research has shown that testwiseness introduced through instructional programs result in improved test scores (Samson, 1985; Dolly & Williams, 1986). The instructional programs have been designed to develop either (a) skills independent of the test constructor or test purpose such as following instructions, monitoring time, and avoiding careless errors (frequently referred in the literature as general test taking skills), (b) skills dependent on the test constructor or test purpose such as cue-use strategies (frequently referred in the literature as testwiseness strategies or skills), or (c) a combination of skills independent and dependent of the test constructor or test purpose.

Samson (1985) performed a meta-analysis covering 24 studies on the effects of training programs for elementary and secondary school children in preparation for achievement tests. The programs varied in length from one to eight weeks or more and consisted of either general test-taking skills alone, or a mix of general test-taking and testwisenss skills such as cue-use. No difference in the mean effect sizes on achievement test performance was found due to the focus of the training program; however, there was a small effect size of .33 favoring elementary and secondary students who participated in the training programs over those who did not. Samson interpreted these results to suggest that the performance of a trained group member at the 50th percentile equaled that of the 63rd percentile of the control group on the achievement test. According to

Samson "training in test-taking skills has a small but significant effect on academic achievement" (p. 262).

Bangert-Drowns, Kulik, & Kulik (1983) examined differences in performance between coached and uncoached groups on various achievement tests given to students from grade 2 to grade 18 (a medical exam), and found similar results to the Samson (1985) study. In the coaching programs reviewed, 21 of the 30 studies contained coaching to develop testwiseness strategies, four studies involved coaching in the content area, 10 of the studies included coaching in anxiety reduction techniques, and 11 permitted practice with test items. The authors attempted to eliminate from the analysis programs devoted to either tutoring in content area or practice by means of taking alternate forms of the test. Overall results revealed a small effect size (.25) in favor of coached students suggesting that coaching, on average, raises scores from the 50th percentile (control group) to the 60th percentile (coached group). More recently, Powers (1993) summarized previous meta-analyses and updated the information with more recent studies on coaching effects on the scores of the Scholastic Aptitude Test (SAT). Although Powers (1993) felt that while standardized test score improvements for college-bound students taking the SAT may not be practically significant, his findings are consistent with those of Samson (1985) and Bangert-Drowns et al. (1983) in concluding that, on average, small positive effects can be expected as a result of coacning. With this study as with the previous studies, there does appear to be a systematic bias in the favor of the coached student suggesting that students benefit from training programs intended to develop skills in one or more of the following areas: (a) general test-taking skills, (b) specific testwiseness strategies, and (c) content area instruction or review.

Although standardized tests may be relatively free of questions where students might benefit from those skills dependent on the test constructor, teacher made tests are not as carefully constructed. Brozo, Schmelzer and Spires (1984) found that 44 % of 1,220 multiple-choice questions from college and university examinations contained testwiseness clues using the Millman et al. (1965) definitions. Brozo et al. reported that 70% percent of the 44% could be answered using the testwiseness clues. Dolly and Williams (1986), using teacher made tests which contained questions where a student could benefit from testwiseness skill, demonstrated statistically significant differences between the higher test scores of students taught testwiseness skills as compared to the scores of the control group. Sarnacki (1979) cited many studies which found significant gains for students receiving any of a variety of training methods.

Measures of Testwiseness

Several researchers (Gibb, 1964; Slakter, Koehler & Hampton, 1970; Diamond & Evans, 1972) have developed tests to examine testwiseness from various perspectives: (a) as a construct,

(b) to examine existence of correlates, and (c) to assess success of training programs. Miller (1990) cited Gibb's (1964) test as being the most comprehensive of the assessment instruments reviewed by Sarnacki (1979).

Gibb's (1964) test of testwiseness measures the use of secondary cues found in test items. Secondary cues in test items can be used to answer the test question itself without content specific knowledge. Although Gibb pointed out that he was well aware that secondary cues are not the only elements which comprise testwiseness, he justified narrowing his focus to cue-using strategies by stating that (a) secondary cues could be effectively examined through at least one type of commonly administered test (multiple choice), and (b) secondary cueing was at least one element of testwiseness which could be controlled for and eliminated as a source of variance by a test constructor should testwiseness be a variable worthy of consideration by examiners.

Gibb's test (1964) was constructed to measure seven specific cue-using skills, with 10 items per skill. However, Gibb did not explore the factor structure with the test as part of his dissertation.

Miller, Fuqua, and Fagley (1990) reported the results of a factor study of the total scores on each of the seven skills from the Gibb Experimental Test of Testwiseness, based on the responses of 181 undergraduates enrolled in "four sections of an upper division educational psychology course required for teacher certification" (p. 205). The authors pointed out that if the number of factors was less than the seven distinct skills suggested by Gibb (1964), then a shorter and more practical version of the test could be devised. Based on a varimax-rotated, principal components solution, Miller et al. reported a two-factor structure for Gibb's test. However, principal components analysis tends to inflate factor loadings over those obtained by common factor analysis, especially when the number of variables is small (Gorsuch, 1990). One reason for this upward bias is that the principal components procedure implies that all the observed variation is common variation and that there is no error variation (e.g., unreliability) associated with any of the variables (Gorsuch, 1983; 1990). Additionally, Miller et al. did not evaluate the stability of the results through cross-validation or hold-out sample, or the bootstrap method (see Efron & Gong, 1983).

The purpose of this paper was to complete a confirmatory factor analysis using common factor loading estimates as a means of evaluating the stability of the two-factor structure of the Gibb (1964) test hypothesized by Miller et al. (1990). If the two-factor model can be confirmed, an equally valid, shorter, more practical test might be devised, facilitating the measurement of testwiseness in future studies or training programs. Since reliability is a problem for any single factor analysis, in that the picture obtained from results of any one factor analysis may change with changes in sample, data collection, and errors of measurement (Hair, Anderson, Tatham, & Black, 1992), a confirmatory analysis was considered an appropriate subsequent procedure for examining the stability of the two-factor structure of the cue-use component of testwiseness reported by Miller et al. (1990).

Method

Subjects

The sample consisted of 173 undergraduate students enrolled in courses in educational psychology and speech pathology classes at two small southern universities. Demographic information was available for 156 of the subjects. Eighty six percent of the students were women, and fourteen percent of the students were men. Of the respondents indicating their ethnicity, 18.6% were African-American, 0.6% were Hispanic, 80.1% were Caucasian, and 0.6% were of other descent. Of the 154 students who reported their age, the mean was 22.2 years with a standard deviation of 5.2. Of the 147 students who reported their grade point average, the mean was 2.98 with a standard deviation of 0.54.

Instrument

The Gibb Experimental Test of Testwiseness (1964) was used to test the cue-use component of testwiseness. The test consists of 70 multiple-choice questions which appear to be difficult history questions, but which can be answered correctly by using cues given within the test question or the test itself. The seven types of cues are: (a) alliterative association cues where a word in the answer is auditorily similar to a word in the question stem, (b) unrelated alternative cues where alternatives to the correct answer are grossly unrelated, (c) specific determiner cues such as "all" or "never" in incorrect responses, (d) precision cues directing attention to a more precise correct alternatives, (e) length cues of obviously longer correct alternatives, (f) grammatical cues such as the correct use of "a" or "an" and correct use of verb tenses, and (g) give away cues where the correct responses to items are given in other test items. These cues are the seven subskills of secondary cue testwiseness that the instrument assesses. Gibb (1964) reported a KR-20 reliability coefficient of .72 for the total score. Miller, Fagley, and Lane (1988) determined a stability coefficient of .64 by administering the test twice over a two-week time period to seventy "junior and senior undergraduate students enrolled in teacher certification courses" (p. 1125).

Though Gibb (1964) did not report a direct appraisal of the validity of his test, total scores on the test were statistically significantly different between a group of undergraduates given training in applying the seven cue-using testwiseness skills and a group not so trained.

Procedure

The test and a demographic questionnaire were completed by volunteer students during one fifty minute class period. Prior to the administration of the instruments, students were informed that the purpose of the project was to investigate test-taking abilities.

Analysis

A confirmatory factor analysis was performed on the Gibb Experimental Test of Testwiseness (Gibb, 1964) using Windows LISREL 8.03 (Joreskog & Sorbom, 1993a) to test the hypothesized two-factor structure of Miller et al. (1990). Additionally, a single-factor model was tested, to determine whether a more parsimonious factor structure--one having a general (cue-using) testwiseness proficiency--would satisfactorily explain the relationships among the seven skill scores.

Pearson product-moment correlations (see Table 1) for the seven skills the Gibb instrument was designed to assess were generated using SPSS (Norusis, 1993). Scores on each skill had a possible range of 0 to 10. Factor loading estimates for the confirmatory analysis were obtained using maximum likelihood, common factor analysis.

Confirmatory Analysis of the Seven Subskills

The Chi-square goodness of fit statistic confirmed that the new data set fit a one-factor model, as well as the two-factor model proposed by Miller et al. (1990), X^2 (14, N = 173) = 15.262, p = 0.360, and X^2 (13, N = 173) = 10.03, p = 0.694, respectively, for the two models. The LISREL 8 estimates using common factor loadings and maximum likelihood factoring procedure (see Table 2) yielded salient factor loadings ($\pm$ .40) for the one-factor model on: Skill 2--Unrelated Alternative Cues (0.51), Skill 4--Precision Cues (0.41), Skill 5--Length Cues (0.48), Skill 6--Grammar Cues (0.58), and

Skill 7--Give Away Cues (0.57). Estimated loadings for Skill 1--Alliterative Cues (0.13) and Skill 3--Specific Determiner Cues (-0.10) seemed to have little or no association with the single factor. For the two-factor model (see Table 3), salient factor loadings for Factor 1 included: Skill 4--Precision Cues (0.44), Skill 5--Length Cues (0.53), and Skill 6--Grammar Cues (0.60), and for Factor 2 were: Skill 2--Unrelated Alternative Cues (0.57), and Skill 7--Give Away Cues (0.66). Again, the loadings for Skill 1--Alliterative Cues (0.15) on factor one, and Skill 3--Specific Determiner Cues (-0.08) on factor two, suggested that these skills are not a part of the two factors, and therefore not strongly associated with the other skills. The summary statistics (Table 1) reveal low mean scores suggesting that responses for test items measuring some of the skills (Skills 1, 3, and 7) were at, or near, guessing levels. The correlation between the two factors in the two-factor model was estimated as 0.725 with a _t_-value of 6.27. The significant _t_-value indicated that the two factors were correlated.

Insert Tables 1, 2 and 3 about here

Other goodness of fit statistics also confirm the one- and two-factor structures. Both Steiger's (1990) Root Mean Square Error of Approximation (RMSEA) and Browne and Cudeck's (1989) Expected Cross-Validation Index (ECVI) as cited by

Jöreskog and Sörbom (1993b), suggest the models fit the data well. For the one-factor model the RMSEA = .0229. Also, according to Browne and Cudeck (1993), cited by Joreskog and Sorbom (1993b), a value of 0.05 or less for RMSEA would suggest a close fit of model and data. The p-value test of close fit (RMSEA < 0.05) = .725 suggested that variations in populations of similar size would not cause a rejection of the two-factor structure. The Root Mean Square Residual (RMSR) = 0.0373. The Expected Cross-Validation Index (ECVI) was less than the saturated model (0.252 < 0.326) indicating that the fitted covariance matrix analyzed fit the model better than an arbitrary model. The goodness of fit statistics suggested an equally good fit for the two-factor model. The RMSEA = 0.0, the p-value test of close fit (RMSEA < 0.05) = 0.913, and the RMSR = 0.0448, all of which suggested a good fit. The ECVI for the two-factor model was less than the saturated model (0.233 < 0.326).

Since the Chi-square statistic is dependent on sample size, alternate indicators which are not dependent upon sample size were examined to determine the fit of the models. According to Joreskog and Sorbom (1993b) the Goodness of Fit Index (GFI) and the Adjusted Goodness of Fit Index (AGFI) "do not depend on sample size explicitly and measure how much better the model fits as compared to no model at all" (p. 122). Both the GFI for the one-factor model (0.975) and the two-factor model (0.984), as well as the AGFI for the one-factor (0.950), and the two-factor model (0.965), indicated a close fit between the model and the data. Indices close to 1 suggest a close fit.

Reanalysis of the Miller et al. Data

Because the present study results showed that, while the Miller et al. (1990) two-factor model did fit the data, a simpler, one-factor model was also satisfactory, a reanalysis of the Miller et al. data was undertaken.

An examination of the data set--Pearson product moment correlations among the seven skill scores--from Miller et al. (1990), confirmed the two-factor model [$\underline{X}^2$ (13, $\underline{N}$ = 181) = 13.511, $\underline{p}$ = 0.409], but not a one-factor model [$\underline{X}^2$ (14, $\underline{N}$ = 181) = 35.54, $\underline{p}$ = .0012]. For the confirmatory analysis of the two-factor model, the LISREL 8 estimates using a maximum likelihood procedure yielded salient factor loadings for all skills except for Skill 1--Alliterative Cues (0.32 on factor 1). The RMSEA was 0.0148, the $\underline{p}$-value of close fit (RMSEA < 0.05) = 0.763, and the RMSR = 0.0445; all suggested a good fit for the two-factor model. The ECVI was less than the saturated model (0.242 < 0.311). The GFI (0.980) and the AGFI (0.957) also indicated a close fit between the model and the data. The correlation between the two factors was estimated as 0.38 with a $\underline{t}$-value of 3.00. The significant $\underline{t}$-value indicated that the two factors are correlated.

For the one-factor model, the RMSEA = 0.0925, the $\underline{p}$-value of close fit (RMSEA < 0.05) = 0.0328 and the RMSR = 0.0751; all revealed that the data fit the model poorly. The ECVI was not less than the saturated model (0.353 > 0.311). Although the GFI (0.943) was acceptable, the AGFI (0.887) was not high enough to suggest a good fit.

Discussion

A confirmatory analysis performed on the seven subskills of The Gibb Experimental Test of Testwiseness (1964) comparing alternative models using LISREL offered partial support of the two-factor structure of Miller et al. (1990). In this study, goodness of fit statistics indicated that a more parsimonious one-factor model fit the data as well as the hypothesized two-factor model. A one-factor model, however, did not fit the data reported in the Miller et al. study. The findings of this study question the identification and stability of the most parsimonious model, and suggest that sample variation appears to affect the stability of the model. Since no means or standard deviations for the skills were reported by Miller et al., it is difficult to ascertain the extent of differences between their sample and that of the present study.

For both the one- and two-factor models in this study, loadings for two of the subskills, Skill 1--Alliterative Cues (0.13 for the one-factor model and 0.15 for the two-factor model) and Skill 3 (-0.10 for the one-factor model and -0.08 for the two-factor model) did not load on either factor suggesting that these skills seem to have little or no association with the factors. All of the other skills yielded salient loadings (see Tables 2 and 3). It should be noted also that the factors were correlated in the two-factor model. For the reanalysis of the Miller et al. (1990) data, the loadings for two-factor confirmatory analysis revealed a loading for Skill 1--Alliterative Cues (0.32) which did not meet the ±.40 criterion set by Miller et al. (1990). All other skills met the criterion for salience. It should be remembered that principal components analysis was used in the Miller et al. study, which tends to inflate loadings. The one-factor model yielded a poor fit for the Miller et al. data, such that four of the seven skills had unacceptable factor loadings [Skill 1--Alliterative Cues (0.27), Skill 2--Unrelated Alternative Cues (0.31), Skill 3--Specific Determiner Cues (0.37), and Skill 7--Give Aways (0.28)].

Miller et al. (1990), making tenative interpretations for the two factor model, suggested that skills loading on Factor 1 which included: 1 (Alliterative Cues), 4 (Precision Cues), 5 (Longer Cues) and 6 (Grammar Cues) seemed to be more overt cues of testwiseness; whereas, the skills loading on Factor 2, which included: 2 (Unrelated Alternative Cues), 3 (Specific Determiner Cues), and 7 (Give Away Cues) seemed to be more subtle cues of testwiseness requiring more "attention or levels of processing of the test questions" (p. 207). The present study, however, did not support the interpretation of Miller et al. (1990). In this study, the highest mean level of performance was observed for detecting the correct answer from among the unrelated alternatives (Skill 2). For this data set the average percent correct by skill was: 41% on detecting the correct answer from among unrelated alternatives (Skill 2), 40% on identifying grammatically correct answers (Skill 6), 39% on identifying the longer and more complete answer (Skill 5), 33% on identifying the most precise answer (Skill 4), 32% on identifying alliteratative cues (Skill 1), 31% on identifying give aways (Skill 7), and 24% on identifying specific determiners (Skill 3). Since performance was best on detecting the correct answer from among unrelated alternatives, it appears that this cue is one of the more obvious cues measured by Gibb's test. The lower percentage correct on locating the correct answer is interpreted as being due to the location of the items near the end of the test, which may reflect a fatigue factor rather than the suggestion that the lower performance is due to a subtle cue requiring more attention.

An alternate interpretation of the two-factor model might be that the first latent dimension or construct represents attention to details related to accuracy evidenced in student's ability to attend to: (a) similarities between words in the stem of the questions and words in the correct response (Skill 1), (b) a more specific response (Skill 4), (c) a longer or more complete or qualified alternative (Skill 5), or (d) a response that is grammatically more accurate than the alternatives (Skill 6). The second latent dimension appears to represent the more obvious cues of detecting unrelated alternatives (Skill 2) and locating give aways (Skill 7). It is possible that, due to the late occurrence of the give away test items, students may have been less diligent due to fatigue, in attempting to locate the answer in an earlier question unless they were strongly motivated or had a persistent nature. If the problem of the late occurences for Skill 7 could be resolved, it is predicted that the percentages of correct anwers for Skill 7 would improve. The "give away" nature of the specific determiner (Skill 3) suggests that it is an obvious cue. It is possible that students do note these cues readily, but perceive them as "trick" questions, and therefore, purposely, and rather consistently, do not respond to the cues.

It should be noted that these results probably reflect the abilities of students who may not have been highly motivated to perform well on the instrument since they were volunteers, and there was no penalty for doing poorly on the test. Fatigue or boredom due to the length of the test may also be a factor in the results. Findings from this study suggest that we may be examining one general factor or construct, test taking skill, or at least aspects of which Millman et al. (1965) referred to as skills dependent of the test constructor or the purpose of the test.

Further research using various populations and sample sizes may indicate the number of factors to which the Gibb (1964) test can be reduced. Gibb, Miller et al. (1990), and the present study used undergraduate participants; using other populations would be helpful in extending this inquiry. Based on this evidence, it appears that a shorter, more practical test which measures fewer subskills than seven might be developed without destroying the validity of the Gibb test. Both the data from the present study and from Miller et al. support a two-factor model as a viable structure for the Gibb test. The stability of the simpler one-factor model requires further investigation.

Since the Gibb Experimental Test of Testwiseness is considered to be one of the best tests for testwiseness (Sarnacki, 1979), continued effort to reduce the length of the 70-item assessment instrument by determining the number of factors, and by interpreting the represented dimensions, merits attention. Providing educators and researchers with a practical and valid test for measuring testwiseness skills can facilitate the evaluation of students or various training programs in an effort to determine if any unfair disadvantage in the testing situation due to low testwiseness skills exists, and to eliminate it.

References

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. C. (1983). Effects of coaching programs on achievement test performance. Review of Educational Research, 53, 571-585.

Brozo, W. G., Schmelzer, R. V, & Spires, H. A. (1984). A study of test-wiseness clues in college and university teacher-made tests with implications for academic assistance centers. (Report No. 84-01). Georgia: College Reading and Learning Assistance Technical Report. (ERIC Document Reproduction Service No. ED 240 928).

Diamond, J. J., & Evans, W. J. (1972). An investigation of the cognitive correlates of test-wiseness. Journal of Educational Measurement, 9, 145-150.

Dolly, J. P., & Williams, K. (1986). Using test-taking strategies to maximize multiple-choice test scores. Educational and Psychological Measurement, 46, 619-625.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. The American Statistician, 37, 36-48.

Fagley, N. S. (1987). Positional response bias in multiple-choice tests of learning: Its relation to testwiseness and guessing strategy. Journal of Educational Psychology, 79, 95-97.

Gibb, B. (1964). Test-wiseness as secondary cue response. (Doctoral disseration, Stanford University) Ann Arbor, MI: University Microfilms, No. 64-7643.

Gorsuch, R. L. (1983). Factor analysis (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Gorsuch, R. L. (1990). Common factor analysis vs. component analysis: Some well and little known facts. Multivariate Behavioral Research, 25, 33-39.

Green, D., & Stewart, O. (1984). Testwiseness: The concept has no clothes. College Student Journal, 18, 416-424.

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1992). Multivariate data analysis (3rd ed.). New York: Macmillan.

Jöreskog, K. G., & Sörbom, D. (1993a). Windows LISREL 8.03. Chicago, IL: Scientific Software International, Inc.

Joreskog, K., & Sörbom, D. (1993b). LISREL 8: Structural equation modeling with the simplis command language. Chicago, IL: Scientific Software International.

Masters, G. N. (1988). Item discrimination: When more is worse. Journal of Educational Measurement, 25, 15-29.

Miller, P. M., Fuqua, D. R., & Fagley, N. S. (1990). Factor structure of the Gibb Experimental Test of Testwiseness. Educational and Psychological Measurement, 50, 203-208.

Miller, P. M., Fagley, N. S., and Lane, D. S., Jr. (1988). Stability of the Gibb (1964) Experimental Test of Testwiseness. Educational and Psychological Measurement, 48, 1123-1127.

Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. Educational and Psychological Measurement, 25, 707-726.

Norusis, M. J. (1993). SPSS for windows: Professional statistics, release 6.0. Chicago, IL: SPSS, Inc.

Powers, D. E. (1993). Coaching for the SAT: Summary of the summaries and an update. Educational Measurement: Issues and Practice, 12, 24-30.

Rogers, W. T., & Bateson, D. J., (1991). Verification of a model of test-taking behavior of high school seniors. Journal of Experimental Education, 59, 331-350.

Samson, G. (1985). Effects of training in test-taking skills on achievement test performance: A quantitative synthesis. Journal of Educational Research, 78, 261-266.

Sarnacki, R. E. (1979). An examination of test-wiseness in the cognitive test domain. Review of Educational Research, 49, 252-279.

Scruggs, T., & Lifson, S. (1985). Current conceptions of test-wiseness: Myths and realities. School Psychology Review, 14, 339-350.

Slakter, M. J., Koehler, R. A., & Hampton, S. H. (1970). Grade level, sex, and selected aspects of test-wiseness. Journal of Educational Measurement, 7, 119-122.

Thorndike, E. L. (1951). Reliability. In E. F. Lindquist (Ed.), Educational measurement (pp. 567-573). Washington, DC: American Council on Education.

Table 1

Experimental Test of Testwiseness Subskill Correlation Matrix,

Means and Standard Deviations (*N* = 173)

	Subskill						
Subskill	1	2	3	4	5	6	7
Subskill 1 Alliterative Cues	1.000						
Subskill 2 Unrelated Alternative Cues	.028	1.000					
Subskill 3 Specific Determiners Cues	.009	-.053	1.000				
Subskill 4 Precision Cues	.040	.149	-.047	1.000			
Subskill 5 Length Cues	.175	.144	-.106	.287	1.000		
Subskill 6 Grammar Cues	.072	.320	-.069	.247	.299	1.000	
Subskill 7 Give Away Cues	.024	.383	-.005	.210	.265	.298	1.000
M	3.15	4.15	2.37	3.32	3.94	3.94	3.21
SD	1.56	1.67	1.30	1.55	1.91	2.09	2.08

Table 2

Experimental Test of Testwiseness Subskills

Confirmatory One-Factor Solution Loadings and t-Statistics N = 173

Subskills	Factor 1
Subskill 1 Alliterative Cues	0.13 (1.34)
Subskill 2 Unrelated Alternative Cues	0.51 (5.58)
Subskill 3 Specific Determiners Cues	-0.10 (-1.09)
Subskill 4 Precision Cues	0.41 (4.47)
Subskill 5 Length Cues	0.48 (5.29)
Subskill 6 Grammar Cues	0.58 (6.39)
Subskill 7 Give Away Cues	0.57 (6.26)

Note: Values in ( ) are *t*-statistics.
Loadings $\geq |.40|$ are considered salient.

Table 3

Experimental Test of Testwiseness Subskills

Confirmatory Two-Factor Solution Loadings and t-Statistics N = 173

Subskills	Factor 1	Factor 2
Subskill 1 Alliterative Cues	0.15 (1.55)	
Subskill 2 Unrelated Alternative Cues		0.57 (5.71)
Subskill 3 Specific Determiners Cues		-0.08 (-0.79)
Subskill 4 Precision Cues	0.44 (4.67)	
Subskill 5 Length Cues	0.53 (5.57)	
Subskill 6 Grammar Cues	0.60 (6.18)	
Subskill 7 Give Away Cues		0.66 (6.25)

Note: Values in ( ) are *t*-statistics.
Loadings ≥ |.40| are considered salient.