DOCUMENT RESUME

ED 214 987 TM 820 243

AUTHOR Rock, Donald A.
TITLE Internal Construct Validity of the Career Skills Assessment Program.
INSTITUTION Educational Testing Service, Princeton, N.J.
SPONS AGENCY College Entrance Examination Board, New York, N.Y.
REPORT NO CEEB-RR-81-10; ETS-RR-81-42
PUB DATE 81
NOTE 22p.; Small print throughout.
AVAILABLE FROM College Board Publication Orders, Box 2815, Princeton, NJ 08541 ($4.00).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Career Development; *Factor Analysis; *Measures (Individuals); *Models; Secondary Education; Test Reliability; *Test Validity.
IDENTIFIERS *Career Skills Assessment Program; *Confirmatory Factor Analysis

ABSTRACT

The primary purpose of this study was to provide evidence for or against the construct validity of the Career Skills Assessment Program (CSAP) instrument. A secondary purpose was to present a systematic procedure for carrying out internal construct validity studies in any testing instrument. Construct validation using confirmatory factor analysis indicated that the CSAP instrument reliably measured what it purported to measure, and five of its six subscales provided sufficiently unique information to make it a useful tool for program and/or individual diagnosis. (Author)

ED214987

College Board Report

No. 81-10

Internal Construct Validity of the Career Skills Assessment Program

Donald A. Rock

Internal Construct Validity of the Career Skills Assessment Program

Donald A. Rock

Educational Testing Service

College Board Report No. 81-10

ETS RR No. 81-42

College Entrance Examination Board, New York, 1981

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in College Board Reports do not necessarily represent official College Board position or policy.

The College Board is a nonprofit membership organization that provides tests and other educational services for students, schools, and colleges. The membership is composed of more than 2,500 colleges, schools, school systems, and education associations. Representatives of the members serve on the Board of Trustees and advisory councils and committees that consider the programs of the College Board and participate in the determination of its policies and activities.

Additional copies of this report may be obtained from College Board Publication Orders, Box 2815, Princeton, New Jersey 08541. The price is $4.

CONTENTS

ABSTRACT

The primary purpose of this study was to provide evidence for or against the construct validity of the Career Skills Assessment Program (CSAP) instrument. A secondary purpose was to present a systematic procedure for carrying out internal construct validity studies in any testing instrument. Construct validation using confirmatory factor analysis indicated that the CSAP instrument reliably measured what it purported to measure, and five of its six subscales provided sufficiently unique information to make it a useful tool for program and/or individual diagnosis.

INTRODUCTION

Construct validation attempts to establish evidence as to whether or not an instrument is measuring what it is purported to measure. Cronbach and Meehl (1955, p. 283) define a construct as "some postulated attribute of people, assumed to be reflected in test performance." Well-known examples of constructs are verbal ability, quantitative ability, and spatial relations. The development of the Career Skills Assessment Program (CSAP) as described in its handbook (College Entrance Examination Board, 1978) postulated a six-construct performance model of career skills. The six performance areas are: (1) self-evaluation and development skills, (2) career-awareness skills, (3) career decision-making skills, (4) employment-seeking skills, (5) work-effectiveness skills, and (6) personal-economics skills. It was not necessarily assumed that all relevant career skills could be subsumed under a six-construct model but that the six areas represented a content core that was common to most developmental guidance programs nationwide, regardless of the specific instructional materials being used.

Since the Career Skills Assessment Program (CSAP) was designed to assess both individual and programmatic progress along the above six dimensions, any internal construct validation must demonstrate that (1) the individual constructs are internally consistent and (2) each measured area furnishes sufficient unique information to be able to make statements about individual and/or program strengths and weaknesses along these dimensions. These two construct validity criteria are critical for formative evaluation. The goal of formative evaluation is both individual and program improvement. It permits feedback to program staff and students on current progress along the measured dimensions, and often results in alteration or reemphasis of the program or instructional sequence.

The goals of demonstrating internal consistency as well as demonstrating differentiation among the measured career skills can best be represented as a problem in convergent and discriminant validity (Campbell and Fiske, 1959). Internal convergent validity suggests that items or subsets of items which were designed to measure a particular career skills construct should "fit" or "load" on their appropriate factor. Given the "fitting" of such a pattern (i.e., a factor loading pattern consistent with the original test specifications), the question then arises, what are the factor intercorrelations given the hypothesized six skill area factor model? Discriminant validity provides some new information about strengths or weaknesses at either the individual or program level.

It is the purpose of this research to marshall evidence for or against the presence (or absence) of the six theoretical factors or constructs which are assumed to underlie the six item content areas. Alternative models which can be hypothesized and tested range from the most parsimonious single-factor model to more complex multiple-factor solutions with varying degrees of complexity in between. Empirical evidence supporting a single-factor model would suggest that the reporting of six separate scale scores would be inappropriate. A fitted single-factor model would argue that the six scales are congeneric measures (Lord and Novick, 1968) of a general career skill. That is, the true scores are perfectly correlated. From a slightly different viewpoint, a single-factor model assumes that the intercorrelations among the observed scale scores corrected for attenuation do not statistically differ from unity. Obviously, there would be no evidence for discriminant validity if the single-factor model held.

Similarly, a factor model of more than one factor but less than six implies that at least one pair of observed scale scores when corrected for attenuation do not statistically differ from unity. Similarly to a single-factor outcome, the reporting of all six separate scale scores would not be justified. Other alternative models might include a specification of more than six factors which include the original hypothesized factors as a subset and/or an entirely different factor structure than that assumed in the original instrument specifications.

To summarize, the alternative construct validity models are: (1) a single-factor career skill model, (2) a reduced space career skill model (more than one factor but less than six), (3) a model of increased complexity which would include the original six constructs embedded in a larger factor model, and (4) an entirely different factor structure from that of the original test specifications. Tests of the above alternative construct validity models can be carried out within the framework of confirmatory factor analysis. If confirmatory factor analysis results are not consistent with models 1 or 2 above but <u>are</u> consistent with the original test specifications, then model 3 would be inconsistent with the empirical findings and model 4 would be superfluous.

That is, with respect to model 4, if the data confirm the original logically-derived specifications there is no justification for trying to "fit" models outside of the original logical framework. However, if the data reject the original construct model, then one must explore other models that may fit the data in hope that a post hoc logic can be developed to explain such data-derived models. However, allowing the data to determine the model is not the classical approach to construct validation and may be more aptly described as collecting and analyzing data in search of a construct.

SAMPLE

A total of 228 juniors and seniors from four different high schools had complete scores on the six career skills assessment scales. The high schools were located in the Northwest and on the East coast. There were 137 juniors and 91 seniors. Although attempts were also made to gather reading scores on these same individuals, the subject who had all six scale scores and a reading measure was much too small to be used in this kind of analysis.

METHOD

Sörbom and Jöreskog's (1976) program for confirmatory factor analysis was used to test (1) a single-factor model and (2) the <u>a priori</u> construct model defined by the original logical framework underlying the six CSAP scales. The six-factor construct validation was carried out under two different specifications. The first specification will be referred to as a "true" or parallel score factor specification and the second will be called a congeneric factor specification. Fitting the <u>a priori</u> model under two different specifications leads to a stronger case for or against the internal convergent and discriminant validity of the CSAP. The "true" score factor specification should approximate an "upper bound" on the internal consistency reliability of the fitted factors while the congeneric factor solution should yield a "lower bound" estimate of the factorially defined scales' internal consistency. If the internal consistency part of the model fits (<u>convergent validity</u>) then the level of intercorrelations among the factors indicat[illegible] information inherent in each factor and thus estimates in a sense the disc[illegible]dity of the factors. If the <u>a priori</u> six-factor model is not inconsistent[illegible]a then the relationships between year in school (junior or senior) and each o[illegible]rs may be assessed. This type of analysis will indicate which, if any, of [illegible] follows a developmental pattern. That is, certain scales may be related to ag[illegible]maturation and/or school curriculum. Appendix A gives a technical discussion of the maximum-likelihood confirmatory factor analysis model used in the various tests of the hypothesized factor patterns.

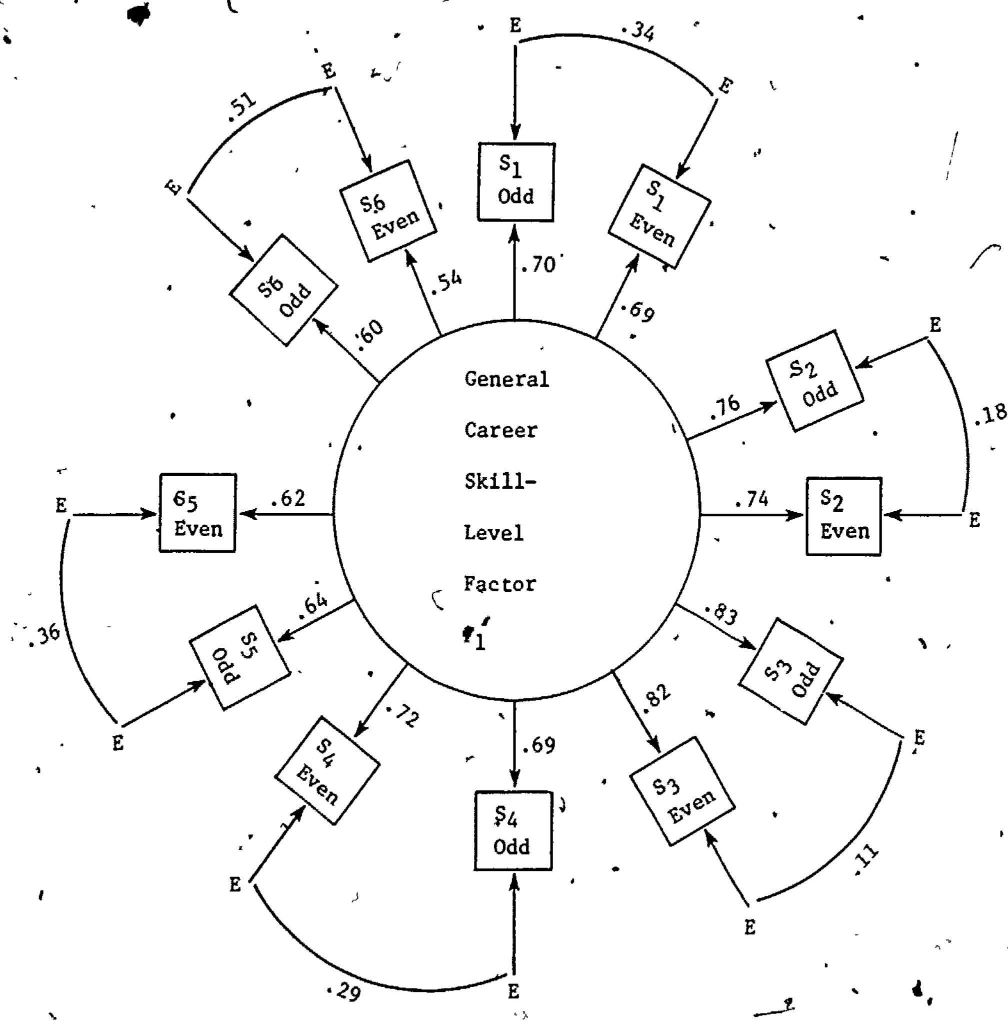

FIGURE 1. Single-Factor Model Solution

S_1 = Self-evaluation and Development; S_2 = Career awareness;
S_3 = Career decision making; S_4 = Employment seeking;
S_5 = Work effectiveness; S_6 = Personal economics.

X^2 with 54 df = 714.68; P = 0.00; X^2/df = 13.23;
root mean square residual = .116.

RESULTS AND DISCUSSION

Single-Factor Model

Figure 1 presents the factor loadings and goodness-of-fit tests for a model which assumes a single factor underlying the CSAP scales. The reader will note that there are 12 rather than six factor loadings.

The 12 variables result from splitting each of the six scales into odd and even item halves. The splitting of each scale into two halves allows for the possible presence of odd-even correlations within scales which are not explained by the single factor. Figure 1 shows that while all odd-even halves representing their respective career-skill areas have statistically significant loadings[1] on the general factor, career awareness and career decision making share the most common variance with the general factor. This would appear to be reasonable since there could be a cause-and-effect relationship between these two skill areas. That is, to a certain extent, appropriate career decision making is conditional on a high level of career awareness.

What is of interest here is the non-zero correlations between the residuals (errors) for corresponding split halves. These correlations are shown in Figure 1 on the curved lines between the odd and even errors. For example, the correlation between residuals for the two indicators of self-evaluation is .34. One can interpret this correlation between any two odd-even residuals as the correlation between odd-even indicators of a particular career-skill area with the influence of the general factor partialled out. If indeed a single factor were the true factor model, these correlated residuals would be close to zero. The probability associated with the large chi-square suggests that the likelihood of observing the sample correlations given a single-factor model is quite low. However, experience has shown that maximum-likelihood ratio tests of goodness of fit tend to reject assumed models even when the residuals are only trivially different from zero. The ratio of the chi-square to its degrees of freedom as well as the root mean square residual[2] are also reported in Figure 1 and will be the primary index of goodness of fit for any given model. Inspection of Figure 1 indicates that the overall root mean square residual $\hat{r}_{ij} = .116$. However, the root mean square residual for just the correlated errors from corresponding odd-even halves within a CSAP skill area was .30, which leads to the judgment of lack of fit of the single-factor model. If the odd-even item clusters were different item types, then these additional skill-area factors might be simply reflecting correlated method variance. Although this is not the case, there remains some danger that these correlated residuals are being generated at least in part by temporary and/or artifactual sources of correlation often observed when correlating odd-even split-halved scores.

Six-Factor A Priori Models

As pointed out before, the test of the *a priori* six-factor model will be carried out under two different specifications. The first and less rigorous test will use the appropriate odd-even pairs to identify each of the six career-skill areas (the so-called true score model). As pointed out in the single-factor model discussion, correlations between odd and even halves are sometimes inflated by temporary sources of covariance and thus could lead to defining factors with little generality across samples or measures. To minimize the possibility of over interpreting such factors, a second congeneric six-factor solution was also tested using skill-area subscales to identify factors. This helps to overcome temporary sources of covariance inflation such as item responses that use the same item stems, etc. This second six-factor model specification will be referred to as a congeneric factor model.

---

1. The maximum-likelihood standard errors of the factor loadings ran from .09 to .10. These standard errors should be interpreted with caution since the correlation matrix was analyzed which, of course, arbitrarily constrains the sample variances to be equal to 1.0.

2. Root mean square residual; $r_{ij} = \left[\Sigma \Sigma r_{ij}^2 / p(p + 1)\right]^{\frac{1}{2}}$; $i \leq j$.

TABLE 1. Six-Factor Confirmatory Model Based on Split-Halved Scale Scores

Split-Halved Scores		F_1	F_2	F_3	F_4	F_5	F_6
Self-eval. and devel.	Odd	.914	0*	0*	0*	0*	0*
Self-eval. and devel.	Even	.899	0*	0*	0*	0*	0*
Career awareness	Odd	0.0*	881	0*	0*	0*	0*
Career awareness	Even	0.0*	840	0*	0*	0*	0*
Career decision making	Odd	0.0*	0*	908	0*	0*	0*
Career decision making	Even	0.0*	0*	975	0*	0*	0*
Employment seeking	Odd	0.0*	0*	0*	875	0*	0*
Employment seeking	Even	0.0*	0*	0*	902	0*	0*
Work effectiveness	Odd	0.0*	0*	0*	0*	879	0*
Work effectiveness	Even	0.0*	0*	0*	0*	862	0*
Personal econ.	Odd	0.0*	0*	0*	0*	0*	943
Personal econ.	Even	0.0*	0*	0*	0*	0*	887

Intercorrelations Among Factors

1.000					
689	1.000				
693	.799	1000			
593	666	464	1000		
445	514	649	612	1000	
322	464	587	487	555	1000

χ^2 with 39 degrees = 64.72
P = .006

χ^2/df = 1.659
Root mean square residual = .027

*Indicates the associated loading was constrained to be zero.

Table 1 presents the results based on the fitting of the "true" score six-factor model. Using the ratio of χ^2 to degrees of freedom criterion, this six-factor constrained solution fits approximately eight times as well as the single-factor model. Using the criterion of root mean square residual, the improvement in fit over the single factor model is by a factor of 4.3. Since the average residual is not practically different from zero and the largest residual is .07, it may be concluded that the assumption of a six-factor model based on the six skill areas is not inconsistent with the observed data.

Although the a priori six factors seem to yield a good fit to the data, the question arises whether a lesser number of factors might also lead to an acceptable fit. That is, is it possible that one or more pairs of the factors defined by the scales are so highly correlated that they could be collapsed into a single factor without a significant increase in lack of fit? Inspection of the intercorrelation matrix among factors in Table 1 indicates that the majority of the correlations are all moderately high but only one, r_{23}, is so high as to lead one to question whether there is indeed two separate factors. The maximum likelihood factor analytic solution yields a standard error for r_{23} of .095 which leads to a p = .05 confidence interval around r_{23} which does not include r_{23} = 1.0. Thus, on a strictly statistical basis it would be inappropriate to collapse factors 2 and 3.

The finding of a high correlation (corrected for attenuation) between career-awareness skills and career decision-making skills is consistent with what was found when the residuals from the single-factor solution were examined. What is encouraging here is that the factor intercorrelations (with the possible exception of r_{23}) are sufficiently low to suggest that the CSAP does meet the criteria for discriminant validity and thus can provide useful information about strengths and weaknesses along at least five of the original postulated dimensions. Similarly, the high factor loadings and the near-zero residuals may be considered evidence for the convergent validity for CSAP.

Stronger evidence for or against the discriminant and convergent validity of the CSAP scales was gathered by fitting the a priori six-factor model using more than two measures to identify each factor and to select these multiple measures so that they minimize the presence of correlated errors. This approach was carried out by using the multiple subscales that define subareas under each of the dimensions as indicators of their respective dimensions. The number of independent career-skill subareas within each of the six dimensions varied from four subscales under career-awareness skills to seven subscales under career decision making, work effectiveness, and personal-economics skills.

The CSAP developers designed these subscales to be independent measures of their respective constructs. Unlike split-halved measures which are likely to approximate parallelism, these independent subscales are only assumed to be congeneric measures of their respective constructs. Therefore, the acceptance of the a priori six-factor model where the factors are defined by the assumed multiple congeneric measures is not only a more powerful test of the presence or absence of the originally specified six factors, but is also a test of the congeneric nature of the independent subscales. That is, this is a more rigorous test of the CSAP construct specification because (1) factors are well over identified since there are more than three indicators of each hypothesized factors, and (2) the factor indicators are independent subscales which are unlikely to suffer from inflated intercorrelations because of the temporary and artifactual sources of overlap that are likely to inflate correlations between split-halved measures.

Table 2 presents the factor loading matrix and intercorrelations among factors based on the six-factor model where each factor is defined by four or more congeneric measures. In spite of the over identification of each factor by the congeneric measures, the goodness-of-fit criteria based on both the χ^2 to df ratio and the root mean square residual suggest quite a reasonable fit. The congeneric model fits almost as well as the "true score" or parallel forms model when the root mean square residual is used as the criterion of fit, and when the χ^2 to df ratio is used the congeneric model appears to fit even better than the true score model. Inspection of the loadings in Table 2 indicate which subscales are the most valid or reliable measures of their particular factor or construct. Almost all the subscales have quite high correlations (factor loadings) with the appropriate constructs. Only one indicator, the sixteenth subscale (the seventh on career decision making) has a correlation in the low fifties ($r = .52$). This subscale was composed of only four items (the smallest scale among the seven) and has to do with taking the "appropriate starting actions." As in the true score model, the factor intercorrelations shown in Table 2 are corrected for attenuation. The congeneric specification of the six-factor model replicates the same pattern of factor intercorrelations found in the true score model. Factors 2 and 3 have the largest correlation, but unlike the other model specification the $p = .05$ confidence interval does contain a correlation of 1.0.

As pointed out earlier, the selection of appropriate career decision-making choices (scale 3) is conditional on career awareness or career knowledge (scale 2). Thus, although the two do not appear to share the same item content, their apparent cause-and-effect relationship does lead to high intercorrelations. In a case such as this, it may be reasonable to report separate scores since "causes and effects" are not logically the same thing, but any interpretation of the differences between these two scores for diagnostic purposes would be questionable. More will be said about this when the factor reliabilities are presented.

TABLE 2. Congeneric Six-Factor Model

	1	2	3	4	5	6
1. Understanding individual differences	0.737	0.0	0.0	0.0	0.0	0.0
2. Evaluating indiv. characteristics and understanding test results	0.809	0.0	0.0	0.0	0.0	0.0
3. Changing personal characteristics and behavior	0.827	0.0	0.0	0.0	0.0	0.0
4. Locating and interpreting information about self	0.790	0.0	0.0	0.0	0.0	0.0
5. Applying knowledge about self to career opportunities	0.809	0.0	0.0	0.0	0.0	0.0
6. Relating abilities, values, needs, and experience to career choices	0.0	0.801	0.0	0.0	0.0	0.0
7. Locating, evaluating, and interpreting information for career choices	0.0	0.769	0.0	0.0	0.0	0.0
8. Knowing facts about career opportunities	0.0	0.845	0.0	0.0	0.0	0.0
9. Finding out about educational requirements for occupations	0.0	0.585	0.0	0.0	0.0	0.0
10. Defining the problem	0.0	0.0	0.593	0.0	0.0	0.0
11. Establishing an action plan	0.0	0.0	0.806	0.0	0.0	0.0
12. Clarifying values	0.0	0.0	0.774	0.0	0.0	0.0
13. Identifying alternatives	0.0	0.0	0.725	0.0	0.0	0.0
14. Discovering probable outcomes	0.0	0.0	0.720	0.0	0.0	0.0
15. Eliminating alternatives systematically	0.0	0.0	0.690	0.0	0.0	0.0
16. Starting action	0.0	0.0	0.518	0.0	0.0	0.0
17. Anticipating job prospects	0.0	0.0	0.0	0.729	0.0	0,0
18. Finding and interpreting facts and sources of information about available jobs.	0.0	0.0	0.0	0.833	0.0	0.0
19. Identifying appropriately written letters, resumes, applications for potential employers	0.0	0.0	0.0	0.827	0.0	0.0
20. Describing appropriate appearance and behavior as one is interviewed and evaluated for job	0.0	0.0	0.0	0.744	0.0	0.0
21. Evaluating specific job in relation to person's needs and interests	0.0	0.0	0.0	0.621	0.0	0.0
22. Identifying responsibilities of employers and employees to each other	0.0	0.0	0.0	0.0	0.751	0.0
23. Developing effective work habits	0.0	0.0	0.0	0.0	0.695	0.0
24. Achieving effective working relationships with co-workers	0.0	0.0	0.0	0.0	0.715	0.0
25. Managing work situations to achieve personal satisfaction	0.0	0.0	0.0	0.0	0.718	0.0
26. Giving and receiving supervision effectively	0.0	0.0	0.0	0.0	0.701	0.0
27. Advancing on the job	0.0	0.0	0.0	0.0	0.637	0.0
28. Planning job changes	0.0	0.0	0.0	0.0	0.633	0.0
29. Figuring your paycheck and income tax	0.0	0.0	0.0	0.0	0.0	0.794
30. Understanding personal banking procedures	0.0	0.0	0.0	0.0	0.0	0.664
31. Purchasing goods and services and paying bills	0.0	0.0	0.0	0.0	0.0	0.813

TABLE 2. Congeneric Six-Factor Model (continued)

	1	2	3	4	5	6
32. Insuring yourself and your possessions	0.0	0.0	0.0	0.0	0.0	0.700
33. Borrowing and using credit	0.0	0.0	0.0	0.0	0.0	0.772
34. Understanding investment procedures	0.0	0.0	0.0	0.0	0.0	0.735
35. Understanding basic economic ideas	0.0	0.0	0.0	0.0	0.0	0.749

Intercorrelations Among Factors

1.000					
0.690	1.000				
0.721	0.808	1.000			
0.610	0.671	0.855	1.000		
0.458	0.526	0.655	0.612	1.000	
0.317	0.454	0.566	0.499	0.551	1.000

$\chi^2/df = 1.318$

Root mean square residual = .056

TABLE 3. Factor Model Reliabilities and Standard Errors of Measurement

Models	Self-Evaluation		Career Awareness		Career Decision Making	
	Rel.	SEM	Rel.	SEM	Rel.	SEM
Parallel (true score)	.90	2.58	.85	2.94	.89	3.00
Congeneric	.90	2.58	.84	3.04	.87	3.26
Grade 11*	.91	3.11	.90	3.21	.91	3.33
Grade 12*	.92	3.12	.88	3.14	.92	3.21
No. of items	60		60		60	

Models	Employment-Seeking Skills		Work Effectiveness		Personal Economics	
	Rel.	SEM	Rel.	SEM	Rel.	SEM
Parallel (true score)	.88	3.12	.86	2.74	.91	2.84
Congeneric	.87	3.25	.87	2.64	.90	3.00
Grade 11*	.91	3.22	.92	2.83	.87	3.32
Grade 12*	.90	3.08	.91	2.83	.90	3.25
No. of items	70		60		60	

*KR-20 Reliabilities and standard errors of measurement as reported in the Career Skills Assessment Program handbook.

Table 3 presents the reliabilities of the scales based on the two hypothesized factor models. Appendix B presents the equations for computing reliabilities of the factors when they are estimated from a confirmatory factor-analysis solution. As pointed out earlier, the congeneric model should yield a lower bound internal consistency estimate since it minimizes the biasing influence of temporary and artifactual sources of covariances. The reliabilities in Table 3 are quite acceptable for all six factorally-defined scales. It is encouraging to note that there appears to be little shrinkage when going from the parallel measures (true score) model to the congeneric model. The lowest reliability under both model specifications is associated with career awareness. This finding is reasonably consistent with the KR-20 estimates which were reported in the Career Skills Assessment Program handbook (College Board, 1978). In general, scale reliabilities based on fitted true-score factor solutions tend to yield the same level of reliabilities as does KR-20 but tend to be slightly lower here since the population of students used in the present study is characterized by consistently smaller subscale variances. Because of the differences in variances, the standard errors of measurement are the more appropriate indicators for comparing accuracy of measurement across populations and estimation techniques. The maximum-likelihood factor analytic estimation of reliability and standard error of measurement does, however, take into consideration information about cross-scale correlations in the estimation procedure. That is, the factor loadings which determine the factor reliabilities are "consistent" estimates since they (the loadings) depend on their interrelationships with all other variables in the factor solution.

The previous finding of the relatively high intercorrelations between career awareness and career decision-making skills along with the somewhat lower factor reliability for career awareness suggest that it might be advantageous to report a single score -- career awareness and decision-making skills--that is, to collapse the two scales and report a five-dimensional career skills pattern rather than the present six-dimensional pattern. This would increase the reliability of the composite scale and also reduce some of the present redundancy.

Table 4 presents the factor extension correlations with grade in school. The correlations are all positive but not significantly different from zero. In this sample, there appears to be little or no relationship between year in school and score level. This lack of relationship may be partly artifactual since there is some evidence (based on the subscale means) that some individuals may have been "topping-out" in this particular sample; that is, the mean scores were within one standard deviation of a perfect score on a number of subscales. This is, of course, consistent with the finding of smaller scale variances.

TABLE 4. Factor Extension Correlations with Grade in School

	Self-Eval.	Career Awareness	Career Decision Making	Employment Seeking	Work Effectiveness	Personal Economics
Grade in School (10th or 11th)	.08	.13	.05	.14	.17	14

The results of this study appear to contradict those of Grandy (1979) who concluded that a single reading factor could explain the correlations between self-evaluation, career awareness, and career decision making. Some of the possible reasons for this apparent paradox are: (1) that their model was not as completely identified since they had only three potential factors, and (2) that the reading factor become quite salient because their sample was characterized by a low reading level. If the latter condition prevailed, score levels on the scales would be conditional on having at least a minimum reading level. Judging from the mean scores in the present sample, the reading levels were probably more than sufficient.

Since Grandy had information on only three subscales, two of which appear to collapse into one factor in the present study, it is not surprising that they came to the conclusion they did. The fact remains that the present study, using a different population and a vastly over-identified model (congeneric model), suggests that there are at least five factors and these five factors are consistent with the original test specifications.

SUMMARY AND CONCLUSION

The purpose of this study was to provide evidence for or against the construct validity of the CSAP instrument. A second purpose was to present a systematic procedure for carrying out internal construct validity studies on any testing instrument.

The Career Skills Assessment Program was designed to assess both individual and programmatic progress along six dimensions. The six areas are: (1) self-evaluation and developmental skills, (2) career-awareness skills, (3) career decision-making skills, (4) employment-seeking skills, (5) work-effectiveness skills, and (6) personal-economics skills. A construct validation plan based on confirmatory factor analytic procedures was implemented that presented evidence for (1) the internal consistency of factors which were fitted according to the original test specifications and (2) the relative independence or uniqueness of five of the content areas. Two of the pre-specified content areas--career awareness and career decision making--appeared to collapse into one factor. Although there was some restriction in score variability in the present sample, the factor reliabilities were sufficiently high to justify the use of five of the CSAP scales as a program diagnostic tool.

In summary, the CSAP appears to reliably measure what it purports to measure, and five of six of its subscales provide sufficiently unique information to make the CSAP a useful tool for program and/or individual diagnosis.

REFERENCES

Campbell, D. T., and Fiske, D. W. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, 56 (1959): 81-105.

College Entrance Examination Board. Implementing the Career Skills Assessment Program -- A Handbook for Effective Program Use. New York, 1978.

Cronbach, L. J., and Meehl, P. E. "Construct Validity in Psychological Tests," Psychological Bulletin, 52 (1955): 281-302.

Grandy, J. An Investigation into the Construct Validity of Three Career Skills Exercises. Research Report 79-9. Princeton, N.J.: Educational Testing Service, 1979.

Lord, F., and Novick, M. Statistical Theories of Mental Test Scores. Reading, Mass.: Addison, Wesley, 1968.

Sörbom, D., and Jöreskog, K. G. COFAMM-Confirmatory Analysis with Model Modification. Chicago, Ill.: National Educational Resources, Inc., 1976.

APPENDIX A: Confirmatory Factor Analysis

Sorbom and Joreskog's (1976) program for confirmatory factor analysis across populations, COFAMM, was used to test various implicit assumptions about the construct validity of the Career Skills Assessment instrument. COFAMM assumes that a factor analysis model holds in each of the g populations under study. In this case, g = 1 since there is only one population under study. If x_g is defined as the vector of the p observed measures in group g, then x_g can be accounted for by k common factors (f_g) and p unique factors (z_g). The model in each population is:

$$x_g = \nu_g + \Lambda_g f_g + z_g \tag{1}$$

where ν_g is a p x 1 vector of location parameters and Λ_g a p x k matrix of factor loadings. It is assumed that z_g and f_g are uncorrelated; the expectation of $z_g = 0$ and the expectation of $f_g = \theta_g$, where θ_g is a k x 1 parameter vector.

Given these assumptions, the mean vector μ_g of the x_g is

$$\mu_g = \nu_g + \Lambda_g \theta_g \tag{2}$$

and the expected variance-covariance matrix Σ_g of x_g is

$$\Sigma_g = \Lambda_g \phi_g \Lambda_g' + \Psi_g \tag{3}$$

where ϕ_g is the variance-covariance matrix of the f_g and Ψ_g is the variance-covariance matrix of z_g. When the factor model does not fit the data perfectly, the observed variance-covariance matrices S_g and observed means will differ from the maximum likelihood estimates of Σ_g and μ_g. The program yields a chi-square statistic that is a measure of these differences, that is, of how well the model fits the data.

The four matrices θ_g, Λ_g, ϕ_g, and Ψ_g are called the pattern matrices. The elements of these matrices are the model parameters, which are of three kinds: (a) fixed parameters, which have been assigned given values, like 0 or 1, (b) constrained parameters, which are unknown but equal to one or more other parameters, and (c) free parameters, which are unknown and not constrained to be equal to any other parameter. A parameter may be constrained to be equal to other parameters in the same and/or different pattern matrices in the same and/or in different groups.

An important feature of a confirmatory analysis is that the parameters of the model may be uniquely estimated, i.e., the model is identified. A solution is unique if all linear transformations of the factors that leave the fixed parameters unchanged also leave the free parameters unchanged. It is difficult in general to give useful conditions which are sufficient for identification. However, at one point in the program the information matrix for the unknown parameters is computed. If this matrix is positive definite, it is almost certain that the model is identified. If this matrix is not positive definite, the program prints a message to this effect, specifying which parameter is probably not identified.

In all tests of the posited CSAP factors the models were over identified, yielding not only unique solutions but sufficient degrees of freedom for statistical tests of goodness of fit. If the model is identified, or over identified as in these examples, standard errors for all the unknown parameter estimates are also provided by the program.

APPENDIX A: Confirmatory Factor Analysis (continued)

In exploratory factor analysis, the model identification often depends on arbitrary restrictions which may have little to do with how the data were gathered. In the confirmatory approach used here, we progressively constrained parameters in equation 3 beginning with the true score model and then the congeneric model. The important point here is that these constraints are applied within the framework of the original test specifications and the nonrejection of each successively stronger model marshals more evidence for the construct validity of the CSAP. Conversely, rejection of the true score model and/or the congeneric model casts doubt on whether the test is measuring the constructs underlying the test specifications.

APPENDIX B: Confirmatory Factor Analysis Estimates of Reliability

I. Reliability of a CSAI subscale

$$r_{ij} = \frac{\Lambda_{ij}^2 \phi_j^2}{\Lambda_{ij}^2 \phi_j^2 + \Psi_{ii}^2} \tag{1}$$

II. Reliability of a factorially defined CSAI scale

$$r_{jj} = \frac{\left(\sum_i \Lambda_{ij}\right)^2 \phi_j^2}{\left(\sum_i \Lambda_{ij}\right)^2 \phi_j^2 + \sum \Psi_i^2} \tag{2}$$

where Λ_{ij} = factor loading of the ith subscale on the jth factor, ϕ_j^2 = variance of the jth factor, and Ψ_{ii}^2 = uniqueness of the ith subscale.

APPENDIX C: Correlation Matrix Among the Six Split-Halved Measures

	1	2	3	4	5	6
1	1.000					
2	0.822	1.000				
3	0.582	0.526	1.000			
4	0.531	0.499	0.740	1.000		
5	0.603	0.570	0.637	0.595	1.000	
6	0.534	0.517	0.615	0.613	0.795	1.000
7	0.407	0.472	0.486	0.496	0.481	0.515
8	0.479	0.550	0.540	0.514	0.530	0.525
9	0.367	0.370	0.370	0.400	0.544	0.498
10	0.331	0.333	0.385	0.399	0.492	0.469
11	0.306	0.281	0.407	0.382	0.498	0.512
12	0.211	0.232	0.341	0.288	0.423	0.428

	7	8	9	10	11	12
7	1.000					
8	0.789	1.000				
9	0.489	0.423	1.000			
10	0.529	0.479	0.757	1.000		
11	0.384	0.399	0.467	0.427	1.000	
12	0.432	0.407	0.448	0.439	0.837	1.000

APPENDIX D: Correlation Matrix Among Subscale Scores

	1	2	3	4	5	6	7	8	9	10
1	1.000									
2	0.607	1.000								
3	0.615	0.648	1.000							
4	0.573	0.627	0.702	1.000						
5	0.595	0.667	0.653	0.621	1.000					
6	0.455	0.494	0.469	0.444	0.480	1.000				
7	0.434	0.412	0.366	0.360	0.368	0.606	1.000			
8	0.421	0.482	0.426	0.494	0.505	0.699	0.630	1.000		
9	0.356	0.300	0.330	0.360	0.316	0.464	0.459	0.495	1.000	
10	0.349	0.363	0.386	0.391	0.342	0.284	0.361	0.401	0.324	1.000
11	0.471	0.534	0.529	0.486	0.559	0.507	0.570	0.577	0.385	0.455
12	0.355	0.417	0.451	0.369	0.477	0.476	0.555	0.519	0.358	0.461
13	0.404	0.437	0.406	0.442	0.489	0.453	0.486	0.499	0.384	0.432
14	0.287	0.349	0.339	0.346	0.365	0.368	0.462	0.462	0.394	0.515
15	0.313	0.435	0.373	0.300	0.322	0.438	0.486	0.424	0.350	0.394
16	0.260	0.376	0.380	0.354	0.396	0.284	0.339	0.361	0.266	0.295
17	0.237	0.299	0.287	0.263	0.324	0.446	0.466	0.439	0.154	0.263
18	0.352	0.425	0.414	0.320	0.448	0.417	0.502	0.461	0.290	0.354
19	0.361	0.485	0.453	0.411	0.530	0.444	0.399	0.429	0.249	0.337
20	0.352	0.416	0.352	0.257	0.450	0.469	0.401	0.442	0.256	0.282
21	0.241	0.359	0.319	0.192	0.327	0.339	0.299	0.342	0.150	0.297
22	0.287	0.311	0.313	0.271	0.298	0.367	0.412	0.373	0.202	0.229
23	0.262	0.268	0.283	0.279	0.312	0.246	0.341	0.345	0.166	0.211
24	0.316	0.318	0.220	0.271	0.230	0.237	0.336	0.304	0.190	0.231
25	0.247	0.187	0.208	0.209	0.211	0.215	0.332	0.316	0.126	0.259
26	0.194	0.208	0.174	0.185	0.196	0.241	0.318	0.302	0.138	0.226
27	0.249	0.275	0.317	0.347	0.317	0.323	0.360	0.317	0.254	0.205
28	0.188	0.224	0.214	0.173	0.258	0.148	0.213	0.230	0.126	0.211
29	0.099	0.232	0.174	0.118	0.247	0.200	0.323	0.305	0.147	0.242
30	0.121	0.099	0.090	0.069	0.140	0.135	0.158	0.201	0.106	0.149
31	0.222	0.248	0.222	0.221	0.291	0.269	0.349	0.420	0.182	0.327
32	0.234	0.326	0.253	0.256	0.305	0.329	0.365	0.420	0.233	0.348
33	0.217	0.213	0.177	0.139	0.258	0.252	0.293	0.351	0.212	0.224
34	0.027	0.071	0.046	0.014	0.146	0.076	0.170	0.196	0.050	0.137
35	0.226	0.258	0.206	0.214	0.285	0.263	0.288	0.343	0.172	0.320

	11	12	13	14	15	16	17	18	19	20
11	1.000									
12	0.652	1.000								
13	0.576	0.512	1.000							
14	0.534	0.582	0.541	1.000						
15	0.509	0.522	0.553	0.532	1.000					
16	0.454	0.429	0.297	0.421	0.333	1.000				
17	0.364	0.387	0.305	0.318	0.347	0.155	1.000			
18	0.440	0.504	0.417	0.431	0.377	0.243	0.613	1.000		
19	0.410	0.388	0.382	0.363	0.353	0.317	0.615	0.693	1.000	
20	0.381	0.413	0.357	0.389	0.332	0.246	0.527	0.588	0.626	1.000
21	0.331	0.327	0.221	0.288	0.258	0.151	0.424	0.508	0.513	0.524
22	0.456	0.378	0.395	0.377	0.452	0.182	0.430	0.461	0.379	0.377
23	0.411	0.369	0.364	0.380	0.339	0.205	0.332	0.370	0.303	0.355
24	0.403	0.291	0.362	0.344	0.332	0.191	0.376	0.354	0.280	0.247
25	0.330	0.306	0.325	0.304	0.324	0.115	0.396	0.416	0.251	0.347

APPENDIX D: Correlation Matrix Among Subscale Scores (continued)

	11	12	13	14	15	16	17	18	19	20
26	0.321	0.329	0.362	0.338	0.350	0.135	0.364	0.398	0.362	0.302
27	0.394	0.341	0.417	0.295	0.357	0.203	0.318	0.359	0.285	0.262
28	0.303	0.330	0.250	0.277	0.337	0.164	0.324	0.352	0.239	0.258
29	0.330	0.364	0.327	0.355	0.410	0.206	0.286	0.425	0.329	0.295
30	0.127	0.206	0.206	0.228	0.229	0.069	0.066	0.225	0.151	0.155
31	0.360	0.427	0.344	0.313	0.335	0.196	0.307	0.438	0.347	0.300
32	0.450	0.463	0.386	0.417	0.415	0.370	0.301	0.377	0.336	0.330
33	0.334	0.344	0.362	0.332	0.379	0.172	0.268	0.374	0.308	0.349
34	0.159	0.201	0.184	0.227	0.264	0.043	0.208	0.230	0.188	0.158
35	0.320	0.345	0.369	0.338	0.382	0.132	0.245	0.342	0.307	0.281

	21	22	23	24	25	26	27	28	29	30
21	1.000									
22	0.291	1.000								
23	0.320	0.502	1.000							
24	0.211	0.560	0.481	1.000						
25	0.274	0.510	0.509	0.549	1.000					
26	0.275	0.494	0.464	0.515	0.547	1.000				
27	0.185	0.483	0.461	0.430	0.415	0.482	1.000			
28	0.215	0.517	0.457	0.437	0.451	0.403	0.406	1.000		
29	0.298	0.367	0.418	0.246	0.323	0.405	0.294	0.331	1.000	
30	0.161	0.149	0.221	0.149	0.221	0.253	0.146	0.206	0.507	1.000
31	0.256	0.290	0.253	0.281	0.395	0.362	0.278	0.271	0.650	0.535
32	0.247	0.362	0.327	0.313	0.277	0.249	0.236	0.245	0.563	0.444
33	0.219	0.269	0.371	0.285	0.279	0.302	0.232	0.263	0.618	0.554
34	0.107	0.153	0.215	0.246	0.270	0.328	0.173	0.245	0.583	0.568
35	0.202	0.295	0.306	0.364	0.333	0.391	0.307	0.367	0.566	0.498

	31	32	33	34	35
31	1.000				
32	0.525	1.000			
33	0.605	0.552	1.000		
34	0.600	0.504	0.593	1.000	
35	0.659	0.523	0.537	0.540	1.000