Perceived and Demonstarted Aptitudes: Same or Different?

by

Dr Poh Sui Hoi
Senior Lecturer
Psychological Studies Division
School of Education, NTU/NIE
469 Bukit Timah Campus
Singapore 259756
Email:POHSH@NIEVAX.NIE.AC.SG

Paper presented at ERA-AARE Joint Conference
on
Educational Research: Building New Partnerships
25-29 November, 1996, Singapore Polytechnic, Singapore

Abstract:

A self-estimate aptitude scale (SEAS) was partly adapted and developed to measure six distinct aptitudes: verbal, numerical, form perception, spatial, clerical and mechanical. The scale was validated for its factor structure using confirmatory factor analytic procedures when administered to a group of 1,568 secondary school students in Singapore schools, together with the same aptitudes from the General Aptitude Test Battery (GATB) and the Differential Aptitude Tests (DAT).

The results from the viewpoint of the Multitrait Multimethod (MTMM) tradition, following Campbell and Fiske (1959) arguments for convergent and discriminant validity did not show that the aptitudes measured by the SEAS, called perceived aptitudes are similar to the aptitudes measured by the standardised tests, called demonstrated aptitudes. Further analysis, using structural equation modelling showed that the perceived aptitudes and demonstrated aptitudes are more likely to be different.

Acknowledgements:

The author would like to thank Prof. P. Dupont of Universite De Sherbrooke, Quebec, Canada for the initial scale which has been adapted and developed further for use in the study. Thanks are also due to Nelson, Canada and the Psychological Corporation for the use of the GATB and DAT in the study. Lastly the author would also like to thank the sponsor, the NIE/NTU, Singapore to this conference and the Ministry of Education, Singapore.

INTRODUCTION

Traditionally, standardized aptitude tests have been used widely, for example in in career counselling. The General Aptitude Test Battery (GATB) and the Differential Aptitude Tests (DAT) are probably the two most well established and widely used aptitude tests in this regard. These are factorially derived multi-aptitude test batteries, consisting of supposedly unifactorial aptitude tests. The use of any of these standardised aptitude test batteries in many instances may not be practicable for administrative reasons -- the long administration time required, the restrictive administration procedures to be followed, and the fact that specially trained personnel are required to administer the test.

A decision was therefore made to develop a Self-Estimate Aptitude Scale (SEAS), which could be considered for use in computerised career guidance packages (or programmes) and for student counselling in schools. The SEAS was planned to be a multidimensional scale measuring a person's perception of his/her aptitudes or abilities. Respondents would be asked to consider and estimate their self-perceptions of the aptitudes linked to those measured by the GATB and the DAT. In the investigation of the construct validity of the SEAS, the factor structure of the scale was examined (See Poh, 1993). Moreover, the SEAS should possess convergent and discriminant validity in an empirical comparison with the tests found in the standardized aptitude test batteries, if indeed they are measuring the same traits.

Theoretical Considerations

Past research in self-estimates of aptitudes are based upon the tacit assumption that the aptitudes measured by standardized aptitude tests and those measured by self-estimates of aptitudes are identical. Hence validation studies almost invariably involved the comparison of self-estimates of aptitudes with their counterparts measured by standardized aptitude tests. The results from the convergent validation studies (eg. O'Hara & Tiedeman, 1959; Tierney & Herman 1973; DeNisi &

Shaw, 1977; Hodgson & Cramer, 1977; Brisco, Muelder & Michael, 1981; Mabe & West, 1982; Lowman & Williams, 1987) showed that the validity coefficients generally were low (<.20s) or at most moderate (about .30s). The conclusions from these studies were that the validity figures, although significantly different from zero in some cases, did not support the proposition that self-estimates of aptitudes can replace aptitudes measured by standardized aptitude tests.

The hypothesis tested in the foregoing research is that the self-estimates of aptitudes and the standardised aptitude tests are measuring identical traits. This hypothesis may not be valid, as suggested by the fact that convergent validity coefficients were low, in many cases not significantly different from zero (eg. Hodgson & Cramer, 1977, Brisco et al., 1981). And, on the face of it, self-estimates seem to be different from measured aptitudes. Self-estimates are measures of the perceptions of one's aptitudes in comparison to one's peers. Here it is assumed that an individual can assess his or her aptitudes, and possesses an internalised standard that can be used to judge this aptitude in relation to others. Measured aptitudes, on the other hand, are the abilities (or aptitudes) one demonstrates in attempting to perform the tasks in a test. The processes involved when one is self-estimating his/her ability seem very different from those required to do well on an aptitude test. Cognitive psychologists (eg. Sternberg, 1977; Sternberg & McNamara, 1985; Embretson, 1985) have argued that the processes involved in responding to spatial aptitude test items are encoding, comparing, searching, rotating, deciding and then responding. It seems evident then, that the processes involved in responding to self-estimate measures of aptitude are less complex than those involved in responding to standardized aptitude test items. One consequence of the difference is that it takes longer to respond to standardized aptitude test items than to the same number of self-estimate items.

Another factor of possible importance in explaining the low correlations between self-estimates and measured aptitudes is that self-estimates, for the most part, have been measured using a single item. This item asks respondents to rate themselves in comparison with others in their own age group on a defined aptitude (trait) using a 5-point or 7-point scale. The generally low validity coefficients when such self-estimate measures are compared to the standardized aptitude tests could be due to the unreliability of the one-item scales. Also, one-item scales lack the checks and balances of multiple items to compensate for the effects of invalidity due to random or other precarious responses.

Purpose of Study

The purpose of present study was to investigate the convergent and discriminant validity of the SEAS in relation to standardized aptitude tests, according to the Campbell and Fiske, 1959 tradition of Multitrait and Multimethod (MTMM) analysis. Modern state of the art procedures involving structural equation modelling will be employed to test the validity of the Campbell and Fiske propositions. Qualitative comparisons will also be made to further ascertain the validity or otherwise of the Campbell and Fiske hypothesis that when two or more methods are used to measure the same construct then they would possess convergent and discriminant validity.

REVIEW OF RELATED LITERATURE

This review is of research into the development and validation of self-estimates (self-ratings, self-reports) of aptitudes or abilities, including a consideration of the designs and methodologies used in the research. First, the concept of aptitude and its measurement are considered, then the concept of validation is discussed.

Aptitude and its Measurement

According to Warren's Dictionary of Psychology (1934), "Aptitude is a condition or set of characteristics regarded as symptomatic of an individual's ability to acquire with training some (usually specified) knowledge, skill or set of responses such as the ability to speak a language, to produce music, etc." This definition of aptitude, although rather old, still holds today. Aptitude is related to the ability or potential of an individual to acquire knowledge or skill of a specified kind. A test of aptitude samples certain response capabilities of the individual at the time of testing, in the expectation that the result will be predictive of what he/she can do in the future. Achievement tests, on the other hand, are designed to measure the effects of a specific program of instruction or training. Achievement and aptitude testing differ in degree of uniformity of relevant antecedent experiences of the examinee (Anastasi, 1990). While achievement tests measure the effects of relatively standardized sets of experiences, such as those provided in a course in oral communication or algebra, aptitude tests measure the cumulative influence of a multiplicity of experiences. A person's aptitudes are thought to be fairly stable, and they can include any characteristic of an individual which predisposes him/her to learning.

Early group tests of mental ability or 'intelligence' often provided a single composite score after sampling responses to tasks from several content domains. This approach was criticised for masking information about the potential of individuals. Anastasi (1990), for example, argued that many of those so called 'intelligence' tests were actually measures of 'verbal comprehension' only. Measurement of aptitudes took a more multi-faceted approach. In the factorially derived General

Aptitude Test Battery (GATB), many skills or aptitudes are tested. Similarly, the Differential Aptitude Tests (DAT) sample behaviours across many skill areas and content domains.

This 'standardized' way of measuring aptitudes requires standard procedures for testing. The batteries of tests are very lengthy and the administration procedures do not permit variations. As career guidance procedures are becoming more 'user friendly,' attempts have been made to measure aptitudes in less structured and less standardized environments.

Self-Estimates of Aptitude

An alternative method for measuring aptitudes is to ask the respondents to estimate their aptitudes for various tasks. This self-estimate approach has been adopted, for example, by Holland (1985) in his Self-Directed Search, which has a subtest dealing with self-estimates of abilities and competencies.

Self-estimates of aptitude (ability) have been shown to have only low to moderate correlations with established aptitude tests (Mabe & West, 1983; Lowman & Williams, 1987). This state of affairs may be due to the nature of the instruments used to elicit self-reports. Dupont (1985) found the instruments in use at that time for self-estimating aptitudes involved only one item per aptitude, an item in which the respondent is asked to rate his/her own aptitude in comparison to the aptitudes of others in his/her age group. Dupont's alternative was to assess aptitudes using several items per aptitude. Each of Dupont's scales consisted of at least 10 items. Subsequently, these scales were adapted for use in the Computerized Heuristic Occupational Information and Career Exploration System (CHOICES) for computer-based career counselling in Canada.

Dupont's items in the CHOICES program were evaluated for their applicability in the Singapore context. It was found that they could be adapted for use, with few modifications. Hence Dupont's scale provided a basis for the development of the SEAS (See Poh, 1993).

Dupont developed her scale according to the aptitudes defined in the GATB. Consequently, she developed nine subscales corresponding to the nine aptitudes purported to be measured by the GATB. For the present study, only six subscales, pertaining to the cognitive as well as the perceptual aptitudes were considered relevant to the purposes for which the SEAS was to be used. The last three of Dupont's aptitudes, each pertaining to a psychomotor skill, were not included in the present work because of lack of contextual relevance and other administrative constraints.

GATB has often been criticised for not including a mechanical reasoning test. Mechanical reasoning has been identified as one of the aptitudes essential for the engineering trades. As an increasing number of Singapore students are going into engineering careers, the SEAS would not be complete without a scale for this aptitude.

The scales of the SEAS can be summarized as follows. Six subscales were adapted from Dupont's scale:

1. General Learning Aptitude (Genapt)
2. Verbal Aptitude (Verapt)
3. Numerical Aptitude (Numapt)
4. Form Perception (Fomapt)
5. Spatial Aptitude (Spaapt) and
6. Clerical Perception (Cleapt)

The Genapt subscale, which corresponds to 'intelligence' in the GATB, actually is made up of verbal, numerical and spatial aptitudes and may not be a specific aptitude, according to the theoretical requirements of the SEAS. The SEAS also has a seventh scale:

7. Mechanical Reasoning (Mecapt).

The Concept of Validation

Validity, according to the Standards for Educational and Psychological Testing (1985), refers to the "appropriateness, meaningfulness and usefulness of the specific inferences made from test scores." (p. 9). From a multiplistic perspective, a variety of inferences may be made from the scores produced by a test, and there are many ways of accumulating evidence to support such inferences. Validity, however, is a unitary concept which refers to the degree to which evidence supports the inferences that are made from the scores (Cronbach, 1971, 1980; Messick, 1975, 1980, 1989).

Traditionally, the various means of accumulating validity evidence have been grouped into categories designated as content-related, criterion-related and construct-related. From the perspective of the prevailing integrated concept of validity, these categories are seen to be convenient facilitators of discourse about validity. Conceptually, however, the categories reflect different aspects of construct validity. Construct validation refers to the process of collecting evidence to support or refute a given interpretation or explanation of test behaviour (Cole & Moss, 1989). Cole and Moss further argued that it is more appropriate to talk about context-based construct validation, as the contextual portion cannot be separated from the construct, and influences all stages of the testing process.

Byrne (1989) argued that "construct validity encompasses two modes of inquiry: (a) the validation of a construct (i.e., trait) and (b) the validation of the measuring instrument." (p. 504). In validating a construct, a researcher seeks evidence in support of hypothesized constructs. In validating a measuring instrument, a researcher seeks evidence that the constructs (traits) purportedly measured by the instrument are the ones actually being measured. Hence in the case of an instrument with several subscales, like the SEAS, evidence of construct validity can be demonstrated if the subscales exhibit a well-defined factor structure that is consistent with the underlying theory.

Related Studies

Research on self-estimates of aptitudes (or abilities) is relatively sparse. A literature search turned up just a few studies, mainly correlational and factor analytic in nature. These have been organized into convergent validity studies and convergent and discriminant validity studies involving multitrait-multimethod (MTMM) matrices. (iv) studies using the cognitive component analysis approach.

Convergent Validity Studies

Mabe and West (1982), reviewed 55 studies in which self-evaluations of abilities were compared with performance tests and other measures (e.g., class grades and supervisors' ratings) of ostensibly the same abilities. Sample sizes in the studies ranged from 40 to 4,300. Product moment correlations were computed for total samples and also for subgroups formed on the basis of sex, race and education. The 267 validity coefficients reported in these studies were generally low to moderate with a mean of .29, but their variability was relatively high (standard deviation = .25). Although the validity coefficients were generally low, Mabe and West argued that these can be improved if the self-raters are trained to make the self-evaluations and are guaranteed that their ratings will not be made public. Also, the self-evaluations in the studies were mainly of the one-item per aptitude type, which could be expected to be relatively unreliable, hence to have correlation coefficients attenuated by measurement errors.

Lowman and Williams (1987) examined the subcomponent of Holland's Self-Directed Search (SDS) involving self-estimates of abilities in six areas. Self-estimates from 149 female undergraduates (aged 18-25) were correlated with objective measures that purport to measure the same abilities. A pattern of low to moderate correlations was found, ranging from .16 to .33. As this subcomponent of the SDS consists of self-estimates on one-item scales, measurement errors might again have substantially attenuated the results.

DeNisi and Shaw (1977) studied 114 students (73 males and 41 females) enrolled in an introductory psychology course. Each subject first rated his/her abilities in 10 areas on a single five point scale per ability. The subjects were then given a battery of tests designed to assess these same abilities. The self-report measure was found to correlate significantly with the corresponding test of ability for nine of the ten ability areas. The Pearson correlation coefficients ranged from .05 for visual pursuit to .41 for numerical ability. DeNisi and Shaw noted, rightly, that the simple product moment correlations obtained in their study may have been attenuated by the small number of rating categories for the items of the self-report instrument. Measurement errors of the type mentioned earlier also would have attenuated the correlation coefficients.

A study by Briscoe, Muelder and Michael (1981) involved 258 high school students of both sexes at four age levels, 15 to 18. Briscoe et al. investigated the relationship between scores on each of the nine ability measures in the GATB and self-estimates by the students of their abilities in each of the same aptitude areas. The self-estimates were made on a single five-point scale per ability. Seventy two validity coefficients were obtained, one for each of the eight independent subsamples. The correlation coefficients ranged from -.02 to .53. Only nine coefficients were statistically significant, eight for females and one for males. Three of the significant coefficients were for Motor Coordination and three others for Manual Dexterity. The authors concluded that with the exception of skills involving psychomotor aptitudes, adolescents, irrespective of age and sex, appeared unable to provide valid self-estimates of their abilities vis-a-vis the validity criterion of a standardized test of the same ability. Again, self-estimates were of the single item type, so the results could be affected by the questionable reliability of the instrument.

O'Hara and Tiedeman (1959), in a study involving 1021 boys in a Catholic High school, found that correlations between the DAT and

self-estimates of the corresponding aptitudes ranged from a low of .03 for Abstract Reasoning by grade 11 students to a high of .58 for Mechanical Reasoning by grade 12 students. Tierney and Herman (1973) studied 1,300 high school students using the same instruments as the previous study. Contrary to the results of O'Hara and Tiedeman, Tierney and Herman found that age, grade level, sex, intelligence and social class of the students had no association with the accuracy of self-estimates of abilities. Again, the self-estimates used in the studies by O'Hara and Tiedeman, and Tierney and Herman were of the one-item type, so the possible effect of unreliability on the correlation coefficients must be acknowledged.

Hodgson and Cramer (1977) investigated the relationship between self-estimates and measures of selected abilities of 74 grade 12 students in an upper middle class suburban school. These students were randomly selected from a group who had taken the DAT four years earlier. The students' self-estimates on Holland's SDS Math, Clerical and Mechanical scales were then compared to scores on subtests of the DAT. A moderate but statistically significant correlation of .41 was found between numerical aptitude and self-estimate of mathematical ability. The correlations for the clerical and mechanical tests of the DAT with the corresponding self-estimates of abilities were small and not significant (-.11 and .19 respectively). Hodgson and Cramer concluded that these results provided little support for the idea that self-estimates of ability by students in high school might replace measurements based upon ability tests. They cautioned, however, that, because the sample size was small, the conclusion was to be held tentatively.

The foregoing convergent validity studies involved correlations between self-estimates of ability and standardized tests of the same-named or similarly named aptitudes. These validation studies were based upon the tacit assumption that the traits measured by the self-estimates and standardized tests are identical. In general, the correlations were found to be low or at most moderate in value, pointing to the conclusion that the self-estimates cannot replace standardized aptitude measures.

Dupont (1985) criticized previous studies of self-estimates of ability on the grounds that they were based on single ratings, hence were likely to be unreliable due to measurement errors. As an alternative, she developed self-estimate aptitude scales based upon several items per aptitude. Each of Dupont's items consists of a statement apparently related to one of the aptitudes measured by the GATB. The respondent is required to indicate on a six-point scale, the extent to which he/she possesses the ability identified in the statement in comparison to others in his/her age group. Dupont's instrument was administered to 402 grade 11 students (226 girls, 176 boys) of average age 14 years, 9 months. Generally, the item-subscale correlation coefficients for the items Dupont created were high (> .60), and internal consistency coefficients (coefficient alpha) were all above .90. Subsequently, Dupont's scale was adapted for use in the CHOICES program for computerised career guidance.

Studies involving Multitrait-Multimethod (MTMM) and Confirmatory Factor Analyses of MTMM Matrices

Campbell and Fiske (1959) proposed the multitrait-multimethod approach for studying the convergent and discriminant validity of a set of tests. They argued that much of validation research is convergent in nature, that is, tests that purport to measure the same trait are

examined to see whether or not they correlate significantly with each other (concurrent validity). Campbell and Fiske went on to note, however, that evidence of discriminant validity is also required, because tests can be invalidated by high correlations with measures from which they were intended to differ. In this kind of research, a multitrait-multimethod (MTMM) matrix is constructed. An MTMM matrix is shown in Figure 1 for illustrative purposes.

Figure 1, about here

Consider Figure 1 in detail: the upper left heterotrait-monomethod triangle contains the coefficient of correlations between different traits measured by method 1. The principal diagonal of this triangle contains the reliability coefficients of measures taken by means of method 1. The lower right heterotrait-monomethod triangle consists of the coefficients of correlation between different traits measured by the method 2, again with the principal diagonal containing the reliability coefficients of measures taken by means of method 2.

The heterotrait-heteromethod rectangle is divided into two triangles, consisting of the correlations between measures of different traits taken by means of different methods. The principal diagonal of this rectangle contains the convergent validity coefficients of correlation between measures of the same trait taken by means of different methods.

Consider the convergent validity diagonal. The coefficients comprising this diagonal should be "significantly different from zero and sufficiently large to encourage further examination of validity" (Campbell & Fiske, 1959, p. 82), thus suggesting the existence of a trait independent of the method of measurement. Next, the reliability coefficients for each method should be sufficiently large to persuade the investigator that some trait is being measured, that the results are not strictly error variance.

Campbell and Fiske specified the following three essential criteria for discriminant validity: First, any coefficient of correlation in the convergent validity diagonal should be higher than the coefficients of correlation lying in the corresponding row and column of the heterotrait-heteromethod rectangle. The basic idea here is that coefficients of correlation involving measures of different methods should be larger when the measures are of the same trait than when they are of different traits. Second, any coefficient of correlation in the convergent validity diagonal should be larger than the heterotrait-monomethod coefficients of correlation involving the same variables as those for the convergent validity coefficient. Here, the reasoning is that, for measures to be valid, there must be more trait variance than method variance, hence measures of the same trait using different methods should correlate higher than measures of different traits using the same method. Third, the pattern of correlations should be the same in all of the heterotrait triangles of both the monomethod and heteromethod blocks.

After Campbell and Fiske's classic work on convergent and discriminant validity, many papers were published describing different ways of analysing the MTMM matrices. Campbell and Fiske had provided no method for quantifying the degree to which their criteria for convergent and discriminant validity were met, hence the judgements to be made were only qualitative in nature. Among the alternatives proposed are the following: application of the analysis of variance in conjunction with

inspection of the MTMM matrices (Kavanagh et al., 1971; King et al., 1980); factor analysis of the MTMM matrices (Golding & Seidman, 1974; Jackson, 1969, 1975); and more recently, confirmatory factor analysis (CFA) of the MTMM matrices (Joreskog, 1968, 1971, 1978), with or without path analysis (Werts & Linn, 1970; Alwin, 1974). The use of CFA has several advantages. Among these are the provision of estimates of parameters, statistical tests of hypotheses and goodness of fit indices (Kenny, 1976; Widaman, 1985; Schmitt & Stults, 1986). The studies reviewed in the following section included the use of confirmatory factor analysis.

Schmitt, Coyle and Saari (1977) analysed a set of MTMM matrices by the following methods: Campbell and Fiske procedure, analysis of variance design, CFA with path analysis (Werts & Linn, 1970), Jackson's multimethod factor analysis, principal components analysis and Tucker's multimode analysis. They found that the CFA with path analysis approach provided the most detailed information regarding individual traits and methods as well as permitting the evaluation of alternative models for the data. They encouraged the use of CFA, where models have to be specified a priori.

Marsh and Hocevar (1983) compared the Campbell and Fiske methodology, the analysis of variance design and the CFA approach in the analysis of MTMM matrix. They listed as the most important advantage offered by the CFA approach, the capability of separating trait, method and unique variances. Marsh (1989), in a later study, cautioned that the CFA of MTMM matrices also encompassed "many problems but few solutions" (p. 335). Among the problems were ill-defined solutions.

The aforementioned analyses show that confirmatory factor analysis of the MTMM matrices offers another approach to the validation of tests. Campbell and Fiske's approach, which is a heuristic approach involving subjective and qualitative interpretations, is still being used despite the criticisms levied at it. In many situations, the CFA of the MTMM adds to and complements, the results obtained by the Campbell and Fiske approach.

METHODOLOGY

Development of the item pool for the SEAS proceeded as follows: For background, a review was made of the literature on self-reports of abilities and tests of abilities, as used in career guidance programs. This included making a search of Tests in Print (Buros, 1974) and the Sixth, Seventh, Eight and Ninth editions of Mental Measurement Yearbooks (Buros, 1965, 1972, 1978; Mitchell, 1985) and reviewing the interest inventories, self-report measures of abilities and aptitude test batteries that were turned up in the search.

The two most well-established and well-researched aptitude test batteries are the General Aptitude Test Battery (GATB) and the Differential Aptitude Tests (DAT). After studying the content and the items of the GATB and the DAT, and also items in the CHOICES computerised career guidance program, items were adapted or written to reflect aptitudes that are deemed important for career guidance in the Singapore context. Most of the items used in the CHOICES program for computerised career guidance in Canada seemed appropriate to the Singapore situation. The developer of these items, Dupont, and the owner of CHOICES, Employment and Immigration Department of Canada, were both contacted, and they gave permission to adapt the items for the present study. The items in CHOICES were used as the basis for the development of the SEAS.

Dupont developed self-report items to match the subtests of the GATB. The GATB covers nine aptitudes; six cognitive and perceptual and three psycho-motor. It was decided that only the six cognitive and perceptual aptitudes would be appropriate for use in a computerized career guidance package (such as JOBS). It was also decided that a mechanical reasoning aptitude, which is not covered in CHOICES, should be created. All the items in the Mechanical Reasoning subscale of the DAT were reviewed, and self-report items were developed to match, as far as possible, the mechanical reasoning that underlies those items. Additional item development was undertaken for the other SEAS subscales, based on the scales in CHOICES. These subscales contained only 10 items. Two additional items were developed, bringing each subscale in the pretest version of the SEAS to 12 items.

To summarize, the seven subscales in the SEAS are as follows:

1. General Learning Aptitude
2. Verbal Aptitude
3. Numerical Aptitude
4. Form Perception
5. Spatial Aptitude
6. Clerical Aptitude
7. Mechanical Aptitude.

Each item of the SEAS consists of a question relating to a certain aptitude, which is defined at the beginning of the subscale for the respondent. The respondent is asked to compare himself/herself to others in his/her own age group, and then rate his/her ability/aptitude on a 5-point scale -- a score of 5 corresponds to a very high rating and a score of 1 corresponds to a very low rating of his/her aptitude.

A draft of the SEAS was sent for comment to experts in Psychology, Guidance and Counselling and Measurement and Evaluation. All the suggestions returned, were taken into consideration in preparing the version of the SEAS that was pilot-tested.

Pilot-Testing

The SEAS was pilot-tested on a group of Singapore Secondary 3 and Secondary 4 students. The main purpose of the pilot-testing was (i) to check whether the wording of the SEAS items was suitable for the intended group, (ii) to see whether or not the students understood the definitions of the aptitudes that were given in the instrument, (iii) to estimate the reliability of the SEAS subscales and (iv) to ascertain whether or not the items in each SEAS subscale were discriminating.

For the pilot test, a typical secondary school in Singapore was selected. (Typical here refers to an average school in terms of mean student performance on a national examination.) The SEAS was administered to a group of Secondary 3 and Secondary 4 students in the Express stream (fast track) and the Normal stream (slow track). The total number of students was 229 -- 120 boys and 109 girls. Of these 105 students were from Secondary 3 and 119 were from Secondary 4, with 108 in the Express stream and 118 in the Normal stream. (These foregoing sets of numbers do not add up to 229 due to missing information for some students).

The pilot-testing was done by the researcher, assisted by the classroom teachers, who were adequately briefed. Students were encouraged to report any problems they encountered with the items. Particular care was taken to explain the definition of each of the SEAS aptitude to the candidates.

The Pilot-Testing Results

As a result of the pilot-testing, it was found that the wording of some items needed to be changed. For example, students pointed out they did not understand the meaning of the word "inventory," which appeared in item 34 in the SEAS. This was later changed to "the quantity of goods or materials in stock." Several female candidates reported that the terms "tune up" (in item SEAS74) and "system of gears" (in item SEAS84) were foreign to them. These two items were not changed, however, as they were deemed to be essential for the mechanical reasoning test and technically sound. The SEAS pilot data were then subjected to an item analysis. The resulting item-subscale correlation coefficients and the Cronbach alpha coefficients are shown in Table 1.

Table 1, about here

The alpha coefficients for all the subscales of the SEAS for the data collected in the pilot study were reasonably high, ranging from .78 to .92, with a median value of .88. The item-subscale inter-correlations, which are taken to be measures of item discrimination, were also reasonably high, ranging from .38 to .74, with a median value of .64 across all seven subscales.

An exploratory factor analysis was carried out on the pretest responses to the 84 item pool. The screeplot suggested that seven factors may be required to explain the common variance of the items comprising the SEAS.

Sampling Design and Subjects

The primary sampling unit for this study was the intact class of students. The classes involved were selected by a two-stage stratification process: First, schools were selected for the sample on the basis of mean student performance for the school on a common national examination. (The examination was the General Certificate of Education 'O Level' Examination, 1989.) The schools were categorized on the basis of these results as high, high medium, low medium and low. Two schools were randomly selected from each category, for a total of eight schools. In the second stage, the classes were selected from the schools on the basis of streams (express and normal). Here the researcher worked with the principals of the schools to arrive at the final sample, making sure that it represented the population as closely as possible in terms of stream and gender of students. The schools and classes were selected in a stratified, random way, but the students were sampled in clusters.

Instruments

Besides the SEAS, the following tests (scales) were administered in the study (definitions of the specific aptitudes were adapted from the GATB Manual, 1983):

1. Verbal Aptitude Test (VerT)

Verbal aptitude is defined as "the ability to understand the meaning of words and to use them effectively, including the ability to understand relationships among the words," and was tested using the corresponding subtest of the GATB. This subtest of the GATB consists of 60 items.

Each item is made up of four words. Respondents are required to identify the two words out of the four for an item that are most nearly the same in meaning or most nearly opposite in meaning. This subtest as reported in the manual has a test-retest reliability of .84. The Hoyt reliability estimates of this test in the present study for the Secondary 3 and Secondary 4 samples were .69 and .82, respectively.

2. Numerical Aptitude Tests (NumT)

Numerical aptitude is defined as "the ability to perform arithmetic operations quickly and accurately." The corresponding subtest of the GATB consists of 50 items. Each item has five options. Respondents are required to work out the answers without the aid of calculators. The answer to each of the items can be obtained using one of the four operations of addition, subtraction, multiplication and division. The reported test-retest reliability for this test was also .84. The Hoyt reliability estimates of this test in the present study for the Secondary 3 and Secondary 4 samples were .89 and .91, respectively.

3. Form Perception Test (FomT)

Form perception is defined as "the ability to perceive pertinent details in objects or pictorial or graphic material and the ability to make visual comparisons and discriminations and see slight differences in shapes and shadings of figures, widths and lengths of lines". This subtest of the GATB consists of four frames of objects, three of which are very nearly the same as the one in the stimulus picture, and one of which is identical to the stimulus picture. Respondents are required to compare the stimulus object with the other four and select the one that is identical to the stimulus object. The reported test-retest reliability was .72. The Hoyt reliability estimates of this test in the present study for the Secondary 3 and Secondary 4 samples were .89 and .93, respectively.

4. Spatial Aptitude Test (SpaT)

Spatial aptitude is defined as "the ability to think of geometric forms in visual terms and recognize relationships resulting from the movement of objects in space". This subtest of the GATB consists of 40 items. Each item consists of a picture (or drawing), with dotted lines where folds are to be made. The four answer options consist of solid objects. Respondents are required to select from the four objects, the one that matches the picture, when folded in along the dotted lines. The reported test-retest reliability was .82. The Hoyt reliability estimates of this test in the present study for the Secondary 3 and Secondary 4 samples were .87 and .89, respectively.

5. Clerical Aptitude Test (CleT)

Clerical aptitude is defined as "the ability to perceive pertinent details in textual material." It is the ability to observe differences in copy, to proofread words and numbers, and to avoid perceptual errors. This subtest of the GATB consists of 150 items. Each item has two names running side by side. Respondents are required to compare the names and indicate whether they are similar or different. The reported test-retest reliability was .77. The Hoyt reliability estimates of this test in the present study for the Secondary 3 and Secondary 4 samples were .96 and .97, respectively.
(Note: the above definitions of the specific aptitudes in quotation marks, were extracted from the Manual for GATB; Nelson, Canada, 1983).

6. Mechanical Reasoning Test (MecT)

Mechanical reasoning is defined as "the ability to understand mechanical principles and devices, and also the laws of everyday physics." (Manual for the DAT; The Psychological Corporation, U.S.A., 1982). This subtest of the DAT consists of 70 items. Each item depicts some mechanical process in pictures. Respondents are required to reason out the process involved and select one out of the three options given that is the correct description of the process. The reported test-retest reliability for boys and girls at grade 9 and 10 averaged .90. The Hoyt reliability estimates of this test in the present study for the Secondary 3 and Secondary 4 samples were .85 and .86, respectively.

Statistical Methods used in Data Analysis

After data collection, the answer sheets were machine-scored and the information was written into files on computer diskettes. The analyses in this study were mainly carried out with the statistical computer programs in SAS, Version 5.18 (SAS Institute Inc., 1985) and the LISREL7 PC Version 7.18 (Joreskog, K. G. & Sorbom, D., 1984-1989).

Analytical Procedures

Several procedures were used in the analysis of the data collected in this study. Each procedure is discussed in turn.

Confirmatory Factor Analysis (CFA)

The underlying factor structure of the SEAS, once shown by the exploratory factor analysis (EFA) procedure to be consistent with the underlying theory, was further validated by confirmatory factor analyses (CFA), using the LISREL program. Joreskog (1969) suggested that a particular application of factor analysis was exploratory or confirmatory depending upon the a priori specification by the researcher. A purely exploratory factor analytic study normally is done to find out how many factors underlie the scores on a set of variables. Once the researcher has ascertained the number of factors present, he/she can then carry out a confirmatory factor analytic study by prespecifying a particular model, consisting of the number of factors and a pre-specified number of parameters to be fixed in value throughout the analysis (eg., fixed at 0) or free to be estimated. The model was then estimated to give the best overall fit to the data, conditional upon the a priori specifications. The first- and second-order factor models tested (See Poh, 1993).

CFA was also applied to analyse the multitrait-multimethod matrices to examine convergent and discriminant validity of the SEAS vis-a-vis the standardized aptitude measures.

Multitrait-Multimethod (MTMM) Analysis

The six distinct subscales of the SEAS purporting to measure the respective aptitudes together with the standardized aptitude tests for these aptitudes were subjected to MTMM analysis. Campbell and Fiske (1959) proposed the multitrait-multimethod approach for efficient examination of convergent and discriminant validity. Here the six traits as well as the two methods were involved. A multitrait-multimethod matrix was then constructed, including all the inter-correlations that resulted when each of the traits was measured by each of the two methods. The convergent validities of the traits were given by the validity diagonal, which consists of correlation coefficients of the heterotrait-monomethod type. Discriminant

validities of the traits were determined by comparing the convergent validity coefficients with the correlation coefficients in the corresponding row and column of the heterotrait-heteromethod block, and the corresponding correlation coefficients in the heterotrait-monomethod triangles involving the same variables. The patterns of the correlation coefficients in all the monomethod triangles as well as the heterotrait-heteromethod triangles were also examined for discriminant validity.

RESULTS

A report of the analyses to determine the convergent and discriminant validity of the SEAS with respect to standardised aptitude tests via Campbell and Fiske postulates and using confirmatory factor analytical procedures were discussed below.

The Six Subscales of the SEAS and their
Relationship with Standardised Aptitude Tests

Scores were obtained for each respondent on each of the six SEAS subscales. Frequency distributions were formed for each of the six SEAS subscales and each of the six aptitude tests. As noted previously, the aptitude tests are here labelled as follows: Verbal Aptitude Test (VerT), Numerical Aptitude Test (NumT), Spatial Aptitude Test (SpaT), Form Perception Test (FomT), Clerical Perception Test (CleT) and Mechanical Reasoning Test (MecT). The distributions for the SEAS subscales symmmetrical and unimodel, as are those for VerT, NumT, SpaT and MecT. The distributions for FomT and CleT are, however, negatively skewed. Other statistics for the variables are provided in Table 2 (Secondary 3 Sample) and Table 3 (Secondary 4 Sample). As can be seen from the Tables, the reliability estimates for all the variables are relatively high, ranging from .69 for VerT to .96 for CleT for the Secondary 3 sample, and from .77 for Cleapt and CE to .97 for CleT for the Secondary 4 sample.

____

Tables 2 & 3, about here

____

The 12 variables were intercorrelated for each sample. The resulting correlation matrices are shown in Tables 4 and 5. For both the samples, four blocks of relationship can be discerned. The upper triangular block indicates that the SEAS subscales correlated positively and substantially with each other. The lower right triangular block indicates that the aptitude tests also correlated positively among themselves, although not to as great an extent as observed for the SEAS subscales. The middle rectangular block generally shows that the SEAS and the aptitude tests are not that highly correlated.

____

Tables 4 & 5, about here

____

The MultiTrait-MultiMethod (MTMM) Analyses

The matrix of intercorrelations among the six SEAS subscales and the six standardised aptitude tests can be viewed as an MTMM matrix (Tables 4 & 5). The rectangular block of correlation coefficients, which is the heterotrait-heteromethod block, with the diagonal coefficients showing

convergent validity contains generally low numbers. Fearing that this could, in part, be due to the skewness of some of the distributions, the FomT and the CleT scales were transformed by the logarithmic transformation and the correlations for the variables were recomputed. The resultant correlation coefficients changed little, however. As an additional check on the validity of the results, scatterplots were made for all the six pairs of variables in the MTMM submatrices in Tables 3 and 4. No obvious non-linear relationships were detected.

According to Campbell and Fiske (1959), convergent validity is established by showing that the correlation coefficient between two measures of the same trait based upon different methods is significantly different from zero and sufficiently large. That is the monotrait-heteromethod correlations must be significant and large. In the case of the Secondary 3 sample in Table 4, it can be seen that significant convergent validity coefficients (in bold) occurred for the numerical (.133), spatial (.149) and the mechanical (.328) traits, but the convergent validity coefficients for the other three traits were not significant. In the case of the Secondary 4 sample (Table 5), significant convergent validity coefficients (in bold) were recorded for verbal (.124), numerical (.239), form perception (.082), spatial (.295) and mechanical (.324) but not for clerical. Even when significant, however, the convergent correlation coefficients were very low. To at most a very limited extent was convergent validity demonstrated for the numerical, spatial and mechanical traits for both the samples. In addition, for the Secondary 4 sample, a small degree of convergent validity was observed for the verbal and form perception traits. The magnitudes of the convergent validity coefficients were not so large, however, as to suggest that the two methods measure the same traits.

Campbell and Fiske defined three criteria for discriminant validity:

1. The convergent validity coefficients should exceed the correlation coefficients between different traits measured by different methods.

2. The convergent validity coefficients should exceed the correlation coefficients between different traits measured by the same method.

3. The pattern of correlation coefficients should be the same within each of the four triangular blocks of the MTMM matrix, the upper left heterotrait-monomethod block, the two off-diagonal triangles in the monotrait-heteromethod block and the lower right heterotrait-monomethod triangle. (Refer to Tables 4 and 5, also Figure 1)

Comparisons by inspection were made across the respective rows and columns of the Secondary 3 and Secondary 4 MTMM matrices of the convergent validity coefficients with the correlation coefficients between different traits measured by different methods (criterion 1) and with the correlation coefficients between different traits measured by the same method (criterion 2). For the Secondary 3 sample, the numerical, spatial and mechanical traits, can be seen to meet the first criterion for discriminant validity, but all the traits failed to satisfy the second criterion. For the Secondary 4 sample, it can be seen that the verbal, numerical, spatial and mechanical traits met the first criterion for discriminant validity, but all the traits again failed to meet the second criterion.

By inspection of Tables 4 and 5, it can be seen that no recurrent pattern was found for the correlations within each of the four triangular blocks. Hence the third criterion of discriminant validity was not met by any of the traits for either sample.

By applying the heuristic procedure of Campbell and Fiske, it can be concluded that the SEAS possesses convergent and discriminant validity to a very limited extent, when compared to standardized aptitude tests.

Confirmatory Factor Analysis of the MTMM Matrices

The MTMM matrix for the two samples was factor analysed to further investigate the convergent and discriminant validity of the SEAS. The hypothesized model specified a priori for the investigation in both the samples was the six-traits and two-methods model, with the traits and methods factors constrained to be uncorrelated as shown in Figure 2. The hypothesis underlying this model is that there are six correlated traits measured by two correlated method factors. This model was compared with alternative models and individual parameter estimates were examined for evidence of convergent and discriminant validity.

____

Figure 2, about here
____

The goodness of fit indices for the series of MTMM models that were tested are reported in Tables 6 and 7 for the Secondary 3 and Secondary 4 samples respectively.

____

Tables 6 & 7, about here
____

Model 1 is the most restrictive model, representing the null hypothesis that each of the observed measure is an independent factor. This null model provides the basis of comparison for the other models. Model 2, the model specified a priori, was supposed to represent the hypothesized relationship among the traits as well as the methods, if it is true that the two methods measure six similar traits. This model, however, produced anomalous results for both the samples. Model 3 is similar to model 2, with the additional specification that the two methods are uncorrelated. This model too produced anomalous results. Model 4 is similar to model 2, with the additional specification that the two methods are equal, thus reducing the number of parameters for method effects from 12 to two. This model again, produced anomalous results. The failure of models 2, 3 and 4 to fit the data suggests that the hypothesized MTMM structure of six traits measured by two methods cannot be confirmed in these data sets. Models 5 and 6 specified six traits, correlated as well as uncorrelated, with no method factors. Here again, for both the samples, the results were anomalous. The failure of these models (5 and 6) to fit the data forces us to reject the hypothesis of six common traits underlying the two sets of measures. Taken together, the five anomalous results lead us to reject the hypothesis that six different traits were measured in common by the two methods. Models 7 and 8 specified no trait factors, only two correlated or uncorrelated method factors. But these two models were identified in the data for both the samples. In this case, model 7 seemed to be slightly better (in terms of fit indices) than model 8. This configuration, shown in Figures 2 and 3 for the Secondary 3 and Secondary 4 samples respectively, can be interpreted as the first set of variables (SEAS subscales) measuring a certain latent trait, and the

second set of variables (aptitude tests) measuring another latent trait. These two different latent traits are correlated to a limited extent.

A detailed examination of model 7 in both the samples revealed that each of the SEAS subscales loaded substantially and significantly on a latent trait specific to these scales, a trait that can be called 'Perceived General Aptitude' (PGA). Similarly, the standardised aptitudes also loaded significantly (but, to a lesser extent) on a latent trait, which can be called 'Demonstrated General Aptitude' (DGA). The unique variances for the measures in both sets were relatively large for both the samples.

____

Figures 3 & 4, about here

____

Summary of Results of Analysis of MTMM Matrices

The Campbell and Fiske approach to the analysis of the MTMM matrices for both the samples, shows that there is slight convergent and discriminant validity in the numerical, spatial and mechanical traits, when the SEAS subscales are compared to the six aptitude test measures. The CFA of the MTMM matrices, however, pointed to the fact that the SEAS and the aptitude tests are better viewed as something different from each other, although what is measured by each set is correlated to a small extent. Clearly, these results provide little support that the self-estimates of aptitudes by students can be equated to aptitudes measured by standardized aptitude tests.

DISCUSSION AND CONCLUSION

The discussion of results here includes an interpretation of the findings together with a consideration of their implications. The limitations of the study and possible directions for future research is also considered.

Results from Multitrait-Multimethod (MTMM) Analyses

The Campbell and Fiske approach to analysing the MTMM matrix, indicated that the SEAS possesses very little in the way of convergent validity when compared to standardised aptitude measures. The magnitudes of the correlations on the SEAS and standardized test measures of an aptitude were low, and did not suggest that one could replace the other.

In an attempt to explain this phenomenon, the items in each of the six standardised aptitude tests were subjected to factor analysis. It was found that each of these tests contained at least two factors for the Secondary 3 as well as the Secondary 4 samples. But each of the subscales of the SEAS was shown to be unidimensional. This could explain, at least partly, why a smaller than expected correlation coefficient was found between each self-estimate aptitude scale and its standardized aptitude test counterpart.

Results from Confirmatory Factor Analysis of the MTMM

To further investigate the convergent and discriminant validity of the SEAS vis-a-vis the standardized aptitude tests, a confirmatory factor

analysis was performed on the MTMM matrix to fit various versions of a model involving six-trait factors and/or two method factors to the data. The results were disappointing as regards the anticipated demonstration of the SEAS scales being matched to standardized aptitude tests. The SEAS and the standardized aptitude tests were best fit by a model in which each set of scales/tests defined a different trait, perceived general aptitude (PGA) and demonstrated general aptitude (DGA). In other words, one's perceptions of one's aptitudes were found in this study to be quite different from objective measures of one's aptitudes. Cognitive psychologists have argued that the processes involved in solving aptitude test items, from the point of view of information-processing techniques, are often complicated and complex (Sternberg, 1985; Embretson, 1985; Pellegrino, 1988). Corresponding processes in responding to self-estimate measures seem likely to be relatively less involved.

This result has significant theoretical implications. In the past, whenever validation studies were carried out, it was assumed that the traits measured by the aptitude tests and those measured by self-estimate instruments should be identical. This assumption seems invalid. Incidentally, Westbrook et al. (1987), in research examining the relationship between scores of measures of cognitive career maturity and self-reported career maturity, found that these two measures were different from each other. Self-reports communicate what individuals are willing to say they believe, feel, have done and can do. What people know, however, may not correspond closely to what they say they know.

Results from Qualitative Analyses of Corresponding Items

To probe further into the apparent lack of convergence between the self-estimate measures and the standardized test scores for each of the six aptitudes, the corresponding sets of items were analysed. The self-estimate items and corresponding standardised test items were compared qualitatively for congruence (The respective definitions of specific aptitudes were extracted from the manual for the GATB, 1983)

Verbal aptitude was defined as "the ability to express oneself in words or in writing." The self-estimate items for this factor were as follows:

SEAS17.Give clear and precise information.
SEAS11.Understand ideas that are presented verbally or otherwise.
SEAS22.Voice your opinions and ideas.
SEAS20.Identify the main ideas in a story.
SEAS23.Talk in a manner that others can understand.
SEAS12.Understand certain explanations.
SEAS13.Understand the meanings of printed words.
SEAS5.Respond sensibly in a discussion.
SEAS19.Explain something to others.
SEAS14.Speak in front of others.

In contrast, the items in the standardized test of verbal aptitude required the students to identify two out of four words that are the SAME or OPPOSITE in meaning. Examples are as follows:

1. a. bigb. largec. dry
d. slow(Answer: a,b)

2.a. dreary b. loyal c. ancient

d. disloyal(Answer: b,d)

3.a. mildb. correctc. wrong
d. similar(Answer: b,c)

4.a. amusingb. tinyc. awkward
d. funny(Answer: a,d)

It is apparent that the standardized test items assess words knowledge. Apart from item SEAS13, none of the Verbal items of the SEAS are related directly to the ability to identify synonyms and antonyms. It is not surprising, therefore, that the correlation between the standardized subtest and the SEAS subscale was low.

In the case of numerical aptitude, which was defined as the "ability to work with numbers quickly and accurately," the standardized test items required the respondent to add, subtract, multiply or divide numbers. But for the SEAS Numapt subscale, the respondent is required to rate himself/herself on the following items:

SEAS27.Add and subtract numbers quickly and
accurately.

SEAS31.Multiply and divide numbers quickly and accurately.

SEAS29.Collect, compare and record numbers accurately.

SEAS28.Figure out how much change to give to someone.

SEAS26.Do mathematical calculations quickly and
accurately.

SEAS30.Calculate the total cost of things you buy.

SEAS35.Count and total up money of different denominations.

SEAS25.Work with numbers.

SEAS36.Take precise measurements of lengths and weights.

SEAS32. Detect errors and mistakes in a page full of numbers.
The items SEAS25, 26, 27, and 31 are related to the four operations tested by the standardized subtest. The differences are therefore not as great as those for verbal trait. But the SEAS scale covers a broader range of numerical skills than the aptitude test, so the fact that the correlation coefficients between the scales were low is, in retrospect, not surprising.

In the case of the trait form perception, which was defined as the "ability to perceive slight differences in shape, match things that are similar and distinguish things that are dissimilar," the contributory items in the SEAS are as follows:

SEAS39. Inspect objects for flaws, grains and texture.

SEAS44.Notice minor scratches or imperfections.

SEAS41. Notice slight differences in thickness of lines.

SEAS42.Distinguish materials that are almost identical.

SEAS38. See tiny things that most people miss.

SEAS47.Imagine what an object would look like when all the pieces are assembled.

SEAS37.See slight differences in shapes in drawings.

SEAS46.Observe differences between microscopic samples.

SEAS45.See differences in facial colour or complexion.

SEAS43. Match things that are similar in shape and colour.

The corresponding items in the standardised test of form perception presented respondents with a stimulus picture followed by four answer options, one of which was identical to the stimulus picture. The respondent was required to pick the identical picture. This kind of task seems relatively simple when compared to those suggested by the SEAS items, with only item SEAS37 showing some resemblance. It is understandable then that the convergent validity figures were low for this trait.

The spatial trait was defined as the "ability to perceive geometrical forms in three-dimensional space." The SEAS items for this trait are as follows:

SEAS60.Draw some furniture.
SEAS51.Create well-balanced drawings.
SEAS59.Make a box from a drawing.
SEAS57.Picture a finished object from a sketch.
SEAS58.Copy a drawing.
SEAS54.Imagine how a box would look like when unfolded.

The typical item of the standardized test of spatial ability provides the respondent with a stimulus picture of a flat piece of metal with dotted lines where the metal is to be bent or rolled in a certain way. The four answer options consist of objects, only one of which can be made by bending or rolling the metal sheet along the dotted line. This task is very similar to that described in item SEAS54. Again, the standardized test appears to have limited overlap with the SEAS scale.

For the trait clerical aptitude, which was defined as the "ability to respond to perceptual tasks routinely," the contributory SEAS items are the following:

SEAS67.Prepare statistical records and reports.
SEAS66.See details in a table of figures.
SEAS64.Use a filing system to sort information.
SEAS68.Take down notes from speeches.
SEAS69.Transcribe notes from tape recorders.
SEAS65.Find hidden words in a text.

In contrast, the standardised subtest on clerical aptitude consists of 150 pairs of names, with respondents being required to indicate within a limited time whether the names in a pair are identical or different. Clearly, such a task is far less involved than those suggested by the SEAS items.

For the last trait, mechanical reasoning was defined as the "ability to understand the mechanical principles and laws of Physics" (Manual, DAT 1982). The significant SEAS items are as follows:

SEAS74.Reason out the direction of movement of certain gears in a system of gears.

SEAS76.Build some simple machines.

SEAS83.Explain the technical features of some machines.

SEAS77.Perceive the underlying pattern of some moving parts.

SEAS75.Understand the working principle of a machine.

SEAS73.Repair broken mechanical toys.

SEAS82.Set up and adjust a machine for certain functions.

SEAS81.Use physical principles to design some new equipment.

SEAS78.Reason out the proper functioning of a machine.

SEAS84.Repair and tune up a bicycle.

With the exception of items SEAS73, 76 and 84, the other seven SEAS items are related directly or indirectly to the mechanical reasoning test. It is not surprising that the convergent validity coefficients for mechanical reasoning were relatively large for both groups of students. (Secondary 3, .33; Secondary 4, .32). These were the highest correlations found between the different measures of a trait. But again, the magnitudes of the correlation coefficients do not suggest that the two methods measure identical characteristics.

The foregoing discussion is an attempt to explain further why low convergent validity coefficients were obtained when the SEAS subscales were correlated with the standardized aptitude tests. Campbell and Fiske had cautioned that in this kind of situation, the experimenter should examine the evidence in favour of several alternative propositions:

i) Neither method is adequate for measuring thetrait,

ii)One of the two methods does not really measure the trait, or

iii)The trait is not a functioning unity.

In the present situation, it could be one or a combination of these propositions that resulted in the low validity coefficients. In fact, it is likely that the SEAS scales and the corresponding standardized tests measure different traits. It can be noted that the failure to achieve significant convergence between the traits measured by the SEAS and their corresponding standardized aptitude measures does not necessarily reflect negatively on either measure. The most that can be argued is that the two measures do not measure the same trait.

The qualitative analysis of the items indicated that the standardized aptitude tests measure a rather restricted set of skills for each aptitude. The SEAS instrument taps self-estimates of a far wider range of skills, and so may preferable for use in career guidance programmes in Singapore.

Limitations of the Study

Despite the efforts to control factors that might jeopardize the validity of the research, certain limitations were unavoidable:

First, the objective standardized aptitude tests were not included in the pilot study.

Second, the data collection required a considerable amount of time for each session. The concentration span of students may have been too short to contend with this many tests.

Third, the present study involved only Secondary 3 and Secondary 4 students in the Express and Normal streams in Singapore schools. Hence there are limitations regarding the extent to which the results can be generalized.

Suggestions for Further Research

The SEAS has been shown to possess sound psychometric properties for the Secondary 3 and Secondary 4 students in Singapore. This scale deserves further investigation to establish its validity across more groups; like the Junior College students and Pre-university students, the Arts students and the Science students; across cultures like Chinese students and Malay students; as well as across occasions to establish its stability and reliability. Studies of this nature will be proposed to the National Institute of Education, Singapore to be carried out on a continuing basis to further establish the validity and reliability of the SEAS. Ultimately, this will improve upon the generalizability of the results. The SEAS should also be validated against career choice in longitudinal studies.

The scale needs to be refined further by adding new items, especially the subscales Spaapt and Cleapt, each of which has only six items. This will be expected to improve the reliability and validity of these scales. Some items in the Verapt subscale, particularly, SEAS5, SEAS11, SEAS13 and SEAS20 need to be rewritten. Several more aptitudes should be considered for inclusion in the SEAS (e.g., the 'entrepreneurial', scientific, and artistic aptitudes). This process of updating and improving on the quality of the instrument would be expected to make the SEAS more useful in the career guidance context, especially as the SEAS will be incorporated in the computerized career guidance program.

Conclusion

The self-estimate aptitude scale (SEAS) was developed parallel to established aptitude tests like the GATB and the DAT. According to the MTMM theory postulated by Campbell and Fiske (1959), similar traits measured by independent methods should show substantial convergent and discriminant validity. Convergent validity studies are conducted on the tacit assumption that self-estimates of aptitudes and measured aptitudes by standardized aptitude tests are identical. In this study, the perceived aptitudes were found not to be closely similar to demonstrated aptitudes. The items in the two types of measures when compared qualitatively, were also found to be very different. The SEAS is a measure of students' self-estimates of their own aptitudes (or abilities) in comparison to their peers. It is a person's perceived aptitude and is not identical to one's demonstrated aptitude. The SEAS has potential for effective use in the career guidance context, and merits further investigation for this purpose.

REFERENCES

Alwin, D. H. (1974). Approaches to the interpretation of relationships in multitrait-multimethod matrix. In H. L. Costner (Ed.), Sociological Methodology, 1973-74; 79-105. San Francisco: Jossey Bass.

American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (1985). Standards for educational and psychological testing. Washington D. C: American Psychological Association.

Anastasi, A. (1990). Psychological testing (6th Ed.). New York: Macmillan Publishing Co. Inc.

Bentler, P. M. & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Briscoe, C. D., Muelder, W. & Michael, W. B. (1981). The concurrent validity of self-estimates of abilities relative to criteria provided by standardised test measures of the same abilities for a sample of high school students eligible for participation in the CETA program. Educational and Psychological Measurement, 41, 1285-1295.

Buros, O. K. (1974). Tests in print II. The Gryphon Press. NJ: Highland Park.

Buros, O. K. (1965). The sixth mental measurement yearbook. Highland Park, NJ: The Gryphon Press.

Buros, O. K. (1972). The seventh mental measurement yearbook. Highland Park, NJ: The Gryphon Press.

Buros, O. K. (1978). The eight mental measurement yearbook. Highland Park, NJ: The Gryphon Press.

Byrne, B. M. (1989). Multigroup comparisons and the assumption of equivalent construct validity across groups: Methodological and substantive issues. Multivariate Behavioral Reasearch, 24 (4), 503-523.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological

Bulletin, 56 (2), 81-105.

Cole, N. S. & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.). Educational Measurement (3rd Ed.). New York: McMillan Publishing Co.

Cronbach, L. J. (1969) Validation of educational measures. Proceedings of the 1969 invitational conference on testing problems: Towards a theory of achievement measurement, 35-52. Princeton, NJ: ETS.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), Educational measurement (2nd Ed.). Washington, D. C: ACJ

Cronbach, L. J. (1980). Validity on parole: How can we go straight? New directions in testing and measurement over a decade. Proceedings of the 1979 ETS invitational conference, 99-108. San Francisco: Jossey Bass.

Cronbach, L. J. (1988). Five perspectives on validation argument. In H. Wainer & H. Braun (Eds.), Test validity, 3-17. Hillsdale, NJ: Lawrence Erlbaum.

DeNisi, A. S. & Shaw, J. B. (1977). Investigation of the uses of self-reports abilities. Journal of Applied Psychology, 62, 641-644.

Dupont, P. (1985). Une mesure d'auto-evaluation des aptitudes profesionelle chez l'adolescent. L'Orientation Professionalle, 21 (1), 47-67.

Fiske, D. W. (1982). Convergent-discriminant validation in measurements and research strategies. In D. Brinberg & L. Kidder (Eds.). New Directions for Methodology of Social and Behavioral Science, 12, 77-92.

Fleishman, J. & Benson, J. (1987). Using Lisrel to evaluate measurement models and scale reliability. Education and Psychological Measurement, 47, 925-939.

Golding, S. L. & Seidman, E. (1974). Analysis of multitrait-multimethod matrices: A two step principle components procedure. Multivariate Behavioral Research, 9, 479-496.

Holland, J. L. (1985). Making vocational choices: A theory of vocational personalities and work environment (2nd Ed.). New Jersey: Prentice Hall Inc.

Hodgson, M. L. & Cramer, S. H. (1977). The relationship between selected self-estimated and measured abilities in adoescents. Measurement and Evaluation in Guidance, 10 (2), 98-105.

Jackson, D. N. (1969). Multimethod factor analysis in the evaluation of convergent and discriminant validity. Psychological Bulletin, 72, 30-49.

Jackson, D. N. (1975). Multimethod factor analysis: A reformulation. Multivariate Behavioral Research, 10, 259-275.

Joreskog, K. G. & Lawley, D. N. (1968). New methods in maximum likelihood factor analysis. British Journal of Mathematical and Statistical Psychology, 21, 85-96.

Joreskog, K. G. (1971). Simultaneous factor analysis in several populations. Psychometrika, 36 (4), 409-426.

Joreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. Psychometrika, 43, 443-477.

Joreskog, K. G. & Sorbom, D. (1989). LISREL VII: Analysis of linear structural relationships by the method of maximum likelihood (Computer program manual). Mooresville, IN: Scientific Software Inc.

Keith, T. Z. (1990). Confirmatory and hierarchical confirmatory analysis of the differential ability scales. Journal of Psychoeducational Assessment, 8, 391-405.

Kelleberg, A. L. & Kluegel, J. R. (1975). Analysis of the multitrait-multimethod matrix: Some limitations and an alternative. Journal of applied Psychology, 60 (1), 1-9.

Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. Journal of Experimental

Social Psychology, 12, 247-252.

Lawley, D. N. & Maxwell, A. E. (1971). Factor analysis as a statistical method (2ed). London: Butterworth.

Lowman, R. L. & Williams, R. E. (1987). Validity of self-ratings of abilities and competencies. Journal of Vocational Behavior, 31, 1-13.

Mabe, P. A. & West, S. G. (1982). Validity of self-evaluation of ability. A review and meta-analysis. Journal of Applied Psychology, 67 (3), 280-296.

Marsh, H. W. & Hocevar, D. (1983). Confirmatory factor analysis of the multitrait-multimethod matrices. Journal of Educational Measurement, 20 (3), 231-248.

Marsh, H. W. (1989). Confirmatory factor analyses of the multitrai-multimethod data: Many problems and a few solutions. Applied Psychological Measurement, 13 (4), 335-361.

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 30 (10), 955-966.

Messick, S. (1980). Test validity and the ethics of assessment. American Psychologist, 35, 1012-1027.

Messick, S. (1989). Validity. In R. L. Linn (Ed.). Educational measurement (3rd Ed.). New York: McMillan Publishing Co.

Mitchell, J. V. (Jr.) (Ed.) (1985). The ninth mental measurement yearbook. Lincoln, Nebraska: The University of Nebraska Press.

O'Hara, R. P. & Tiedeman, D. V. (1959). Vocational self concept in adolescence. Journal of Counselling Psychology, 6 (4), 292-301.

Poh, S. H. (1993). Investigating the validity of a self-estimate aptitude scale (SEAS) for secondary school guidance in Singapore. Unpublished PhD thesis, University of Toronto, Canada.

Ray, M. L. & Heeler, R. M. (1975). Analysis techniques for exploratory use of the multitrait-multimethod matrix. Educational and Psychological Measurement, 35, 255-265.

Schmitt, N. Coyle, B. W. & Saari, B. B. (1977). A review and critique of analyses of multitrait-multimethod matrices. Multivariate Behavioral Research, 12, 447-478.

Schmitt, N. & Stults, D. M. (1986). Methodological review: Analysis of multitrait-multimethod matrices. Applied Psychological Measurement, 10 (1), 1-22.

Sternberg, R. J. (1977). Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities. Hillsdale, N. J.: Lawrence Erlbaum Associates.

Sternberg, R. J. & McNamara, T. P. (1985). The representation and processing of information in real-time verbal comprehension. In S. E. Embretson (Ed.). Test design: Developments in psychology and psychometrics. Florida: Academic Press Inc.

Tierney, R. J. & Herman, A. (1973). Self-estimates ability in

adolescence. Journal of Counseling Psychology, 20 (4), 298-302.

Werts, C. E. & Linn, R. L. (1970). Path analysis: Psychological examples. Psychological Bulletin, 67, 199-212.

Westbrooke, B. W, Stanford, E. E. & Donnelly, M. H. (1990). The relationship between career maturity test scores and appropriateness of career choices: A replication. Journal of Vocational Behaviour, 36, 20-32.