TRAINABILITY ASSESSMENTS AND WORK SAMPLES:

A FIELD STUDY AND A META-ANALYSIS.

A thesis presented in partial fulfillment of

the requirements for the degree of

Master of Arts in Psychology

at Massey University

Karl PAJO

1987

This Thesis is Dedicated

to Faye and Karl

ABSTRACT

This study compared a work sample test with a trainability test for the prediction of typing students grades. A meta-analysis of the work sample literature was also carried out. Participants in the work sample trainability test comparison were 89 female first year Polytechnic typing students. Students were randomly assigned to either the work sample group or the trainability test group. Tutors then administered the relevant predictor and data was collected. Scores on the predictors were later correlated with the students grade in their second terms test. All the obtained correlations were found to be highly significant although the results unexpectedly revealed that the error score on the work sample was the best predictor overall. It was suggested that the tutors inexperience in administering trainability tests, their greater familarity with work samples and certain deficiencies in the criterion may have contributed to the unexpected trend in the data. Meta-analysis was used to cumulate and average results from many different studies which examined work samples. Studies which utilised training criteria were analysed seperately from those which employed job proficiency criteria. Results from the analysis showed substantial remaining variance following correction for statistical artifacts. The studies were then grouped according to Robertson and Kandola's (1982) classification of work samples in order to identify potential moderator effects. Meta-analysis of subgroups revealed that for all categories, with the exception of group discussion/decision making, considerable variance still remained following correction for statistical artifacts. It is suggested in the discussion that further research on work samples is required, particularly the development of a classificatory system which can accurately and reliably distinguish between types of work samples. Possibilities for future research on trainability tests are also explored.

ACKNOWLEDGEMENTS

I wish to thank my supervisor, Dr Mike Smith for his encouragement, advice and patience throughout this study. I would also like to express my gratitude to Dave George who was always willing to provide assistance. My thanks also go to Dr John Spicer who helped resolve several knotty statistical problems.

I am grateful to the staff of the Manawatu Polytechnic who freely gave of their time and facilities and enabled this study to take place.

Finally, it is with deep appreciation that I acknowledge the support of my parents and all my friends. Larry and Sharon who often had to put up with me until the early hours of the morning. Mike, Julie, Steven (Thumis), Fiona, Steven (Sudagura), Chris, Mark, Wendy, Yvonne, Scott, Josie, Sharon, and Gerald with whom many good times were had. Also my ex-flatmates, Mark, Leeanne, Joanne and Jin who tolerated my many foibles.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

Page

LIST OF FIGURES

Page

FIGURE

CHAPTER ONE.

INTRODUCTION

Personnel psychology, which constitutes the psychology of personnel decisions (Landy, 1985), has always been a traditional area of interest for many industrial and organisational psychologists. As early as 1917 tests were being developed to predict the success of employees on the job. Both the first and second world wars provided further impetus for the testing movement as a whole and allowed personnel psychologists the opportunity to develop and refine their skills (Grant, 1980). The use of tests and other selection procedures burgeoned and they became increasingly sophisticated.

Hakel (1986) points out in his review of personnel selection that there has been substantial progress in the last few decades. He notes that in the early 1960's personnel selection could best be considered "pragmatic, empirical, and atheoretical." Since then personnel research has diversified and other facets of selection have begun to receive some of the attention they merit. Tenopyr and Oeltjen (1982) note that there was a long overdue upsurge in research on job analysis. Hakel's (1986) review indicates that this research continues unabated. The measurement of performance, particularly the cognitive processes underlying performance assessments, has been a topic that has also generated a substantial body of literature. Utility analysis (e.g. Hunter and Hunter, 1984; Schmidt, Hunter, McKenzie, & Muldrow, 1979) has developed to the point where psychologists can safely describe the monetary savings to be gained from the implementation of efficacious selection procedures. The requirement in many countries that selection should not adversely affect minority groups has also prompted a significant amount of work on the part of numerous psychologists. Meta-analytic procedures have now been developed and offer a comprehensive means for assessing cumulative results. These developments have been paralleled by the continuing refinement of existing predictors and the development of alternative selection strategies.

Today the personnel practitioner has a considerable array of selection tools available from which to make a choice. While the techniques available have multiplied, the fundamental goal of personnel selection has remained unchanged. The primary objective is still the identification of those applicants who best suit the organisation's requirements. This usually involves the prediction of an applicant's likelihood of success on the job or during training. Unfortunately, despite continuing research and development, the reliability and validity of many predictors can hardly be considered reassuring.

Work samples however, offer a promising alternative selection strategy which has been receiving increasing attention from psychologists and others interested in matching people to jobs. Historically, research on work samples has tended to be rather sparse and fragmentary. Downs (Note 1) has attributed this to several factors. The zeitgeist of the time was one that favoured the development of predictors that conformed with traditional psychometric test properties. There was an implicit belief amongst psychologists that the predictor should be different from the criterion (Wernimont & Campbell, 1968). Psychological tests were to be used as indicators of predispositions to behave in certain ways rather than being regarded as examples of the typical behaviour of individuals (Robertson & Kandola, 1982). This notion was contested when Wernimont and Campbell (1968) argued that validity would be enhanced if predictors resembled more closely the criterion behaviour they were attempting to predict. They described this concept as behavioural consistency and it is what Asher and Sciarrino (1974) have labelled point-to-point correspondence between predictor and criterion space.

Other more practical difficulties also tended to preclude the widespread acceptance of work samples when used for the prediction of competency on the job. Work samples are disadvantaged in that they have to be individually designed and validated for each particular job. Furthermore, they are often costly to set up, particularly if complex machinery or simulations are required, and are also expensive in terms of manpower and materials used. In contrast paper-and-pencil tests can be used off the shelf, are generally easier to administer and cheaper. Given that such tests were considered capable of measuring those abilities important to performance on the job it is not surprising that work samples were regarded with suspicion (Downs, Note 1).

A further bar to the ready acceptance of work samples as predictors was their past history of use in organisations. Work samples had traditionally been associated primarily with the verification of the acquisition of skills (Downs, Note 1). Essentially their role was to function as an achievement test assessing competence and providing a basis for certification. In some cases they were also used as a criterion to validate other selection methods.

Thus, the practical problems in developing and administering work samples, their past history of use predominantly as criteria to which organisations and individuals had become accustomed, and the adherence to traditional psychometric test concepts, all conspired to inhibit the propagation of work samples as a selection tool. However, proponents of work samples refused to be deterred and continued with a great deal of enthusiasm to investigate this potentially useful predictor. Their work generated a great deal of interest and a considerable body of literature (see reviews by Asher & Sciarrino, 1974; Howard, 1985; Robertson & Kandola, 1982; Downs, Note 1; Gill, 1979; Karren, 1980; Robertson & Downs, 1979; Hunter & Hunter, 1984).

There are many reasons for this upsurge in research on work samples, albeit that it has been somewhat sporadic at times. The gradual alteration of the belief that predictors and criterion should be different has been alluded to. Tests were used to sample behaviour based on the premise that the best indicator of future performance is past performance. Wernimont and Campbell (1968) and Asher and Sciarrino (1974) argued that prediction of future behaviour would be facilitated if tests more closely resembled the behaviour to be predicted (the criterion). Thus, the notion of behavioural consistency or point-to-point correspondence was established. The development of work samples was also given a boost by the failure on the part of more established selection procedures to reach acceptable levels of validity despite in some cases, several decades of research and experimentation.

One prominent example of a commonly used selection device is the employment interview. There have been several reviews of its reliability and validity. One of the earliest was that carried out by Wagner (1949). He reported a median validity coefficient of .27 for the 22 studies reviewed. Reilly and Chao (1982) summarised the data from 12 studies and came up with a mean validity coefficient of only .19. Such disappointing results have since become typical of research in the field (e.g. Mayfield, 1964; Ulrich & Trumbo, 1965; Wright, 1969; Schmitt, 1976). Recent reviews (Arvey, 1979; Arvey & Campion, 1982; Reilly & Chao, 1982) have also pointed to the interviews susceptibility to bias and distortion and particularly the fact that it may act as a vehicle for discrimination against women and minority group members. All in all, the literature suggests that the interview may not be an efficacious method for selecting personnel.

Similar conclusions can be reached regarding the use of references. Reference reports are commonly requested by many organisations (Muchinsky, 1979). The veracity of reference reports however, is questionable and their widespread use difficult to justify. In a review of the available literature Muchinsky (1979) concludes that reported validity coefficients ranged from unacceptable to mediocre. Reilly and Chao (1982) report an average validity coefficient for the studies they reviewed of .14. It appears that reference reports are unlikely to contribute appreciably to the validity of employee selection decisions.

Psychological tests are commonly used for personnel selection in occupational settings. Ghiselli (1973) reviewed the validity of aptitude tests during the period 1920 to 1971. Tests were classified into broad categories. These included;

a) tests of intellectual ability

b) tests of spatial and mechanical ability

c) tests of perceptual accuracy

d) tests of motor ability

e) tests evaluating personality and/or interests.

The average predictive validity for each category of test was then calculated for different occupational groupings. Average validity coefficients for both training and proficiency criteria rarely exceeded .30. However, as Ghiselli (1973) notes, various artifacts such as restriction of range in predictor and criterion scores, errors in sampling, and unreliability of predictor and criterion measures would mean that such estimates are likely to be conservative hence underrating the true predictive power of the tests.

A more recent review incorporating meta-analytic procedures capable of correcting for such artifacts was conducted by Hunter and Hunter (1984). Using formulas developed by Hunter, Schmidt, and Jackson (1982) they were able to correct the variance across different studies for sampling error and wherever possible also corrected for the effects of error of measurement and range restriction. Ghiselli's (1973) review of ability tests was reanalysed using these more sophisticated procedures. The conclusions reached by the authors was that most of the variance in results across studies was due to sampling error. Furthermore, the validity figures they computed were markedly higher than those obtained by Ghiselli (1973). The average validity of cognitive ability tests for different job families ranged from .27 to .61. The average validity for tests of psychomotor ability ranged from .17 to .44. Multiple correlations computed using combined cognitive and psychomotor ability scores tended to be uniformly high across all the job families. Excluding the job of sales clerk, the validity for the combined tests of ability ranged from .43 to .62. Hunter and Hunter (1984) determined that the average validity of cognitive and psychomotor ability tests combined was .53.

In addition to Ghiselli's (1973) study Hunter and Hunter (1984) reanalysed the data from several other reviews. Relevant figures are presented in table one.

TABLE 1 META-ANALYSIS DERIVED FROM (A) DUNNETTE (1972) (B) REILLY AND CHAO (1982).

PREDICTORS	No OF CORRELATIONS	AVERAGE VALIDITY
A Cognitive Ability	215	.45
Perceptual Ability	97	.34
Psychomotor Ability	95	.35
Biographical Inventories	115	.34
Interviews	30	.16
Education	15	.00
Job Knowledge	296	.51
Job Tryout	20	.44
B Biographical Inventory	44	.38
Interview	11	.23
Expert Recommendation	16	.21
Reference Check	7	.17
Academic Achievement	10	.17
Self Assessment	7	some
Projective Tests	5	little
Handwriting Analysis	3	none

from Hunter and Hunter (1984)

The table shows that in general, most selection instruments are poor predictors. The major exceptions appear to be tests of ability, biographical inventories, and work samples.

The use of biographical data appears to be a promising approach to selection. However, there are shortcomings associated with its use. Empirically keyed biodata scores are prone to attenuation of validity over time (e.g. Wernimont, 1962; Hunter & Hunter, 1984). Furthermore, unless a cross validated research design is used, the process of deriving a biographical inventory is one that is prone to massive capitalization on chance (Hunter & Hunter, 1984). It is also possible that applicants may intentionally falsify their responses (e.g. Goldstein, 1971).

A straightforward appraisal of the value of work samples has been impeded by the failure on the part of researchers to come to a clear agreement about what actually constitutes a work sample. Depending on the text consulted, one's impression of a work sample may differ substantially (e.g. Cronbach, 1966; Guion, 1965). This confusion in the literature has persisted (e.g. Howard, 1983; Landy, 1985) although attempts at a rapprochement have been made (Downs, Note 1; McCormick & Tiffin, 1976; Thornton & Byham, 1982). The basis for most disagreement has centreed on how broad or narrow the definition of a work sample should be. Downs (Note 1) abstracted what she considered were the key features of work samples commonly agreed upon by most writers in the field. Using those features she derived a definition of a work sample test as

> "a performance test based on work or job related elements, the design of which allows for measurement or objective assessment of the skills involved in all, or crucial aspects of the job. This measurement may be used to measure past learning or predict potential to learn in the future." (page 2)

Such a definition is quite broad in scope and would include a variety of tasks or tests that vary along a dimension of "fidelity" or relatedness to actual work performance. Examples of work samples could thus range from business games, in-basket tests, leaderless group discussions, through to trainability tests, job simulations and measurement of performance at the job station. Acceptance of such a definition would go a long way towards clearing up many of the misunderstandings currently rife in the literature and would set the field on a theoretically sounder basis.

Useful distinctions within the domain of work samples can still be made. For example, Asher and Sciarrino (1974) classify work samples as either motor or verbal. A motor work sample is a task involving the manipulation of things (e.g. performance on an aircraft simulator, piecing together an electronic circuit board). A verbal work sample is a task containing problems which are primarily people or language oriented (e.g. an in-basket test, leaderless group discussion). Their review demonstrated that motor work samples were superior in predictive power to all other predictors except for biographical data. The verbal work sample tended to be consistently less efficient in its ability to forecast job proficiency than the motor work sample but was still superior to most of the other predictors. When the relevant criterion was changed to "success in training" then the verbal work sample was clearly superior to the motor work sample.

Robertson and Kandola (1982) differentiate between four categories of work sample;

1) Psychomotor. Tasks involving the physical manipulation of objects.

2) Individual, situational decision making. Tasks in which the applicant is required to make decisions similar to those made in the job being tested for. This category can vary along a dimension of realism with close approximations being in-basket tests while more abstract cases could involve the presentation of hypothetical situations and asking the applicant how he/she would respond.

3) Job-related information. Typically paper-and-pencil tests, their purpose is to evaluate applicant knowledge in areas considered to be directly relevant to work performance.

4) Group discussions/decision making. A group of individuals are required to discuss a particular topic and their performance during the discussion is assessed.

They found that psychomotor work samples and job-related information tests had the highest median validity coefficients (.39 and .40 respectively) and the greatest proportion of coefficients above .40. Situational decision making was the poorest of the four categories with the lowest median validity coefficient (.28), the greatest proportion of coefficients below .30 and the smallest proportion above .40. Comparison to other psychological tests showed that psychomotor work samples were superior to all other types except for biographical data.

Group discussion measures also produced quite high validity coefficients in comparison with other tests. An interesting feature of Robertson and Kandola's (1982) analysis is that the high validities obtained by job-related information tests seemed to be mainly confined to situations where training criteria were used. When one considers only the criteria of job performance then the median validity coefficients for psychomotor, group discussion and situational decision making work sample tests outstripped those of job-related information tests.

Hunter and Hunter (1984) compared a number of alternative predictors of job performance using meta-analytic procedures. Abstracting data from many studies, including other meta-analyses, and using the criterion of job performance, as measured by supervisor ratings, they compared predictors used for entry level jobs where training followed hiring and predictors used for decisions regarding promotion or certification. Work samples were second only to an ability composite in predictive power for entry level jobs (mean validity of .44). They were the most efficient predictor used for promotion or certification decisions (mean validity of .54).

Gordon and Kleiman (1976) have conducted one of the few studies which has directly compared a work sample with a standardised test. They used recruits from three separate classes that attended a police training academy. The training program was a 20 week course during which recruits were instructed in the fundamentals of police work. Recruits were administered a work sample after approximately two weeks on the course which covered areas such as introduction to law enforcement, the relationship of the police department to other civic agencies, department rules and regulations and organisation of the department. The recruits were also administered a standardised intelligence test. Correlations with the trainability criterion (sum of the grades achieved during the training course) revealed that in all cases the work sample achieved significant validity coefficients (ranged from .52 to .72) whereas only one of the validity coefficients for the intelligence test was significant (range from .15 to .56).

Mount, Muchinsky, and Hanser (1977) compared the predictive and concurrent validity of a work sample with two traditional paper-and-pencil tests under controlled laboratory conditions. The work sample consisted of following a diagram and constructing a model from mechanical parts. The criterion was the assembling of a more complex model. Using the number of parts correctly assembled as the dependent measure the authors found that in all cases the validity of the work sample was higher than that of the paper-and-pencil tests. Furthermore, even when all three predictors were combined using multiple regression there was only a slight improvement in the validity coefficient obtained over and above that of the work sample alone.

Similar results have been obtained by Sylvia Downs in her work on trainability assessments. An early study (Downs, 1970) involved the development of a trainability test for sewing machinists in a children's clothing factory. The company's existing selection procedures (a form board and a pin board) were compared with the trainability test. The trainability test was highly predictive of success at the end of training while the other selection procedures failed to achieve any significant predictive validity. The results were so convincing that the company immediately terminated its old selection methods and embraced the new test whole-heartedly.

Smith (1977) compared university entrance examination marks, the mechanical aptitude test and the space relations test from the Differential Aptitude Test Battery and a specially designed trainability test for the prediction of the practical performance of dental students. Using students from three separate academic years he found that the trainability assessment was highly correlated with performance on a combined criterion of conservation test marks and final conservation exam marks. In fact, the trainability test surpassed all other predictors except for the DAT mechanical reasoning test.

Siegal and Bergman (1975) constructed trainability tests (what they called miniaturised job training and evaluation) and compared them with standard US Navy paper-and-pencil tests for the prediction of performance by low aptitude naval recruits. Scores on six trainability assessments and three navy tests were correlated with judges' ratings of recruits' performance on several job related tasks after nine months' fleet experience and after 18 months' fleet experience. For the first follow-up, five of the six job performance criteria were predicted better by the trainability tests than by the navy selection tests. For the second follow-up, some attenuation of predictive power for the trainability tests was apparent and the navy predictors were superior for five of the six criterion tasks. Siegal and Bergman (1975) note that such attenuation is not unusual and may simply reflect that the trainability tests are more appropriate for predicting success on initial job entry rather than subsequent improvement. Cohen and Penner (1976) have expressed some reservations regarding the methodology of Siegel and Bergman's (1975) study, particularly the failure to cross-validate the predictors and the large number of drop-outs in the sample used.

Other authors have used trainability tests and although they were not compared with alternative predictors, high validity coefficients have been reported for such diverse jobs as carpentry (Robertson & Mindel, 1980), welding (Downs, 1968; cited in Robertson & Downs, 1979; Robertson & Mindel, 1980), fork truck operating (Downs, 1972), electronic assembling (Smith, 1972), industrial sewing (Downs, 1972), metal use and fitting (Smith & Downs, 1975), brick laying, capstan operating and centre lathe turning (Robertson & Mindel, 1980), catering and forestry work (figures reported in Downs, Note 1), and naval recruits training to be firemen, seamen and airmen (Siegal, 1983).

Changes in beliefs about the functions work samples can fulfil and the clear demonstration in many studies of their superiority over other predictors coupled with the high validity coefficients attained have served to popularise work samples as a viable selection procedure. Concomitant with this was a rise in interest about fairer selection spurred on in many cases by legal changes and social pressures requiring that tests should not exhibit adverse impact. Researchers interested in work samples were able to capitalize on this since in many cases traditional selection procedures were proving to unfairly discriminate against women and/or ethnic minorities (see Arvey, 1979; Einhorn & Bass, 1971). It was argued that work samples would not be prone to such effects since nothing could be fairer than selecting an applicant based upon his or her performance on a sample of the work he or she would actually be required to do. Several studies have subsequently confirmed that work samples do appear to be a fair method of selection.

Schmidt, Greenthal, Hunter, Berner, and Seaton (1977) compared a work sample for metal trades skills with a well constructed content-valid, written achievement test for the same technical area. The written achievement test and each of its component subtests showed large and significant minority-majority differences. The work sample showed a considerably smaller difference between minority-majority workers. Schmidt et al (1977) explain that the small gap exhibited was primarily due to differences in work speed and suggest that this minimal amount of adverse impact could be reduced by decreasing the weighting of the work speed sub-score in the work sample. The authors also point out that both minority and majority examinees saw the job sample tests as significantly fairer, clearer, and more appropriate in difficulty level.

Hamner, Kim, Baird, and Bigoness (1974) conducted a laboratory study in which they examined the way the sex and race of the rater and the sex and race of the ratee influence assessments of ratee performance on a simulated work sampling task. Their results suggested that sex-race stereotypes do influence assessments of behaviour on a work sampling task although unexpectedly the ratings of women's performance were inflated rather than deflated. Brugnoli, Campion, and Basen (1979) criticised the research of Hamner et al (1974) on the grounds that the work sample selected failed to represent important performance factors and hence may have encouraged raters to rely on stereotypes when evaluating applicants. They also argue that Hamner et al (1974) should have used an evaluation device more specific to the behaviours being observed rather than a global rating scale and non-behavioural anchors. They then designed an experiment to examine the role of evaluation specificity and task relevance in explaining racial bias in the use of work samples. They found that bias was not evident when subjects used behavioural recording forms or when evaluations were based on observations of relevant job behaviour. They conclude that if work samples are carefully developed and raters focus on and record relevant behaviour then the potential for bias in the use of work samples appears small.

Bray and Howard (1983; cited in Howard, 1983) report large racial differences when the paper-and-pencil School and College Ability Test was used. Use of an in-basket exercise showed considerably less adverse impact, and performance in group discussions showed almost none. Cascio and Phillips (1979) constructed motor and verbal work sample tests for use by a US city government. No significant difference in selection rates for minority versus majority workers was reported. Downs (1970) found no significant difference in the trainability assessment ratings and criterion ratings of United Kingdom applicants and overseas applicants for the job of sewing machinist. She also notes that there was a high degree of agreement between ratings of overseas applicants on the trainability test and their criterion performance ratings . However, separate validity data for the two groups was not presented and, as Robertson and Kandola (1982) point out, studies that fail to report validity data are only of limited use. Some exceptions are studies by Grant and Bray (1970), Field, Bayley, and Bayley (1977), and Kesselman and Lopez (1979) who all report improved validity accompanying the use of appropriate work sample tests and reduced adverse impact for minority groups (see Robertson and Kandola, 1982 for a review).

Favourable applicant reaction and other ancillary functions of work samples have also contributed to their increased usage. Downs (1970) reports that both instructors and applicants preferred a trainability test over existing selection procedures. Instructors liked the test because they felt more involved in the selection procedure. Applicants liked the test because they felt it was fair and enabled them to demonstrate their capabilities. Schmidt et al (1977) reported that both minority and majority subjects in their study considered the job sample test as significantly fairer, clearer, and more appropriate in difficulty level than a written test covering the same content area.

There is also some evidence that work samples could function as realistic job previews. Wanous (1977) in a review of realistic job previews has concluded that they allow applicants to make more informed choices hence diminishing subsequent dissatisfaction and increasing the probability that applicants will remain on the job. While not all the literature is consistent with such a view (e.g. see Reilly, Brown, Blood & Malatesta, 1981), studies using work samples do seem to offer some support. Downs, Farr & Colbeck (1978) examined the data from sewing machinist trainability tests administered throughout the United Kingdom during the period 1973-1975. All applicants who sat the test were invited to start work regardless of the grade received. The authors found that the individual's trainability assessment grade (ranging from A-highest to E-lowest) influenced the decision about whether or not to start work. Fully 90.8% of those graded A accepted the companies offer while 81.1% of those graded B, 75.6% of those graded C, 54.6% of those graded D and only 23.1% of those graded E accepted offers of employment. The evidence suggests that the trainability tests allowed applicants to accurately gauge their own performance and encouraged self-selection based on those judgments.

Farr, O'Leary & Bartlett (1973) found that for white subjects the administration of a pre-employment work sample resulted in more accurate expectancies about task requirements and a commensurately lower voluntary turn-over rate. The failure to find similar results for black subjects was explained in terms of the differential importance of factors in the work situation. It was argued that black applicants may have paid more attention to such facets of the environment as pay and interpersonal relations whereas whites may have focussed exclusively on the task related factors portrayed in the work sample.

Additional support for the notion that work samples may encourage the self assessment of ability comes from a study by Downs (note 2). She administered a trainability assessment to Royal Navy helicopter pilots. Results of the assessment showed that pilots' ratings of their own abilities were clearly affected by the trainability test. She concluded that the test helped applicants to judge whether or not they would like the job and enabled those who did well to assess themselves more realistically.

Campion (1972) asserts that work samples may have additional advantages of reducing the possibility of response sets and faking and being less prone to charges of invasion of privacy. While such a claim seems inherently plausible the paucity of relevant data in the literature means that such statements remain to be substantiated.

The use of work samples is not completely without drawbacks. Several authors (e.g. Downs, Note 1; Howard, 1983; Robertson & Downs, 1979; Smith & Downs, 1975) have enumerated their disadvantages. These include the fact that work samples (particularly psychomotor work samples, job-related information tests and trainability tests) tend to be job specific. This means that they have to be individually designed and validated for different jobs. Furthermore, they require continual monitoring in order to ensure that their reliability and validity is not affected by changes in job content over time. In cases where machinery is required it can be costly to set up or to construct appropriate simulations. Many work samples can only be administered individually or in small groups and require skilled assessors to evaluate performance. They usually take longer to administer and also use more materials than equivalent paper-and-pencil selection tests. Howard (1983) also notes that work samples may not be particularly useful for assessing a candidates' range of knowledge. Finally, some studies (Downs, 1977; Siegal & Bergman, 1975; Smith & Downs, 1975) suggest that the predictive validity of work samples may be prone to attenuation over time (perhaps due to changes in job content as noted above). However, as Hunter and Hunter (1984) point out, there exist very few predictors which do not become less efficient with the passage of time.

The present study elaborates on previous research examining the value of work samples. More specifically it is composed of two parts. The first involves a direct comparison of the predictive validity of a work sample and a trainability test. Robertson and Downs (1979) distinguish between standard work samples and trainability tests. A trainability test is a specialised type of work sample designed to evaluate an applicant's potential to **learn** a task or to succeed in **training**. Such tests typically include standardised instructions and a period of demonstration during which the instructor teaches the applicant the task. While the applicant is being instructed in the task he or she is permitted to ask questions and to practice. The applicant is then tested on the material he or she has been taught by being asked to perform the task unaided. The applicant's performance on the task is assessed by the instructor who uses a standardised error checklist and rating scale. Thus the trainability test differs from the normal work sample in several important ways.

1) it incorporates a structured learning period during which the applicant is encouraged to ask questions and practice the task.

2) the assessor uses an error checklist rather than simply evaluating the product of performance on the work sample.

3) the applicant is only tested on what he or she has been taught during the learning period, hence, it does not assume any prior experience.

Research on trainability tests has shown that they are very good predictors of success in training and often subsequent performance on the job (see Robertson & Downs, 1979 for a review). The question remains as to whether or not trainability tests tap important performance dimensions that work samples do not. Trainability tests tend to be more complex and time consuming than equivalent work samples. Employers may be reluctant to accept trainability tests on face value unless it can be clearly demonstrated that they are superior to work samples for predicting training outcomes. A study by Gordon and Kleiman (1976) found that performance on a work sample administered to police recruits was significantly related to grades achieved at the end of training. In other words, Gordon and Kleiman (1976) were able to predict trainability using a work sample. While such evidence is suggestive it is by no means conclusive. There has been no study to date that has specifically compared a work sample with a trainability test.

Statement of Hypotheses.

For the first part of the present study it is hypothesised that a trainability test designed to predict training success for typing students will prove to be superior to a work sample administered for the same purpose. Such a hypothesis is based upon the fact that trainability tests are specifically designed to forecast training outcomes and their prior history of success in that endeavour.

The second part of the study is a partial replication and extension of work done by Robertson and Kandola (1982). It consists of an examination of the predictive validity of different types of work samples. Robertson and Kandolas' (1982) categorization of work samples will be used with the addition of a separate trainability test group. More sophisticated meta-analytic formulas will be used to analyse the data rather than simply calculating distributions of validity coefficients and median validity coefficients.

CHAPTER TWO.

METHOD

SUBJECTS

Subjects for the comparison of the trainability test and the work sample consisted of 89 female students enrolled in an introductory typing course at the Manawatu Polytechnic. Subjects differed in the amount of prior typing experience. The majority of subjects had done a small amount of typing at school with three to four years being most common. A few had some work experience and approximately one quarter of the class were novice typists. Subjects ranged in age from fifteen years to twenty-two years with the majority falling into the fifteen and sixteen year age bracket. Subjects were randomly allocated to either the work sample (N=47) or to the trainability test group (N=42).

PROCEDURE

Initially, the course supervisor was interviewed and a basic text on typing consulted in order to develop a list of those skills students should be competent in by the end of the training course. This information was then used to derive a task analysis questionnaire (see appendix B) which could be used as the basis for the construction of an appropriate trainability test and work sample. The typing tutors were asked to rate how important each of 57 factors were in distinguishing a well trained typist from a poorly trained typist. They were then required to identify and rank what they considered to be the fifteen most important tasks. An identical questionnaire was sent out to ten students who had participated in the previous year's course and were still resident in Palmerston North. The students were asked to rate the same 57 tasks according to degree of difficulty of learning. After completing the rating they were asked to identify and rank the fifteen most important tasks. The information from tutors and students was combined to ensure that during the development of the trainability test and work sample important elements of the typist's job were included. Once the tests had been developed they were submitted to the course supervisor and the typing tutors who assessed their suitability. Once the tests had been approved the tutors were given practice in their administration and scoring. The tests were then administered to the students during normal class hours and relevant data was collected. Trainability test and work sample scores were then correlated with grades obtained during their second terms test (see appendices C and D for a copy of the procedural flow chart used by the tutors who administered the trainability test and the work sample). It was felt that the student's mark in the terms test was a more rigorous and objective criterion measure than the more traditional supervisors ratings. The test required no interpretive marking and if ratings had been employed the opportunity for criterion contamination would have been much greater since it was the supervisors who administered the predictors. The terms test was also the most convenient criterion available in that it interferred least with the duties of tutors and the progress of student learning. Chadwick-Jones, Nicholson, and Brown (1982) have pointed out that occupational research in applied settings is subject to many constraints. They argue, scientific rigorousness notwithstanding, that researchers must often make do with what data can be collected since more desirable measures may well disrupt the work routine and incur additional expenses for the organisation.

Trainability Tests.

Downs (1984) has described trainability tests as a type of performance test in which the applicant is instructed about how to make a single work piece or how to operate part of a process and is then systematically assessed whilst performing the task under controlled conditions.

There a five main steps to be followed when constructing a trainability test.

1) Job Analysis: This is conducted to identify relevant operations and accompanying skills required for successful performance on-the-job.

2) Selecting the Work Piece: The workpiece used in the trainability test should fulfil certain criteria. It should be based on crucial elements of the job and should be sufficiently difficult so that a range of observable errors can be made. On the other hand it should only incorporate those skills which are able to be taught during the instruction period and should not be excessively long so that

applicants are unable to finish. Most trainability tests typically require twenty to thirty minutes to complete.

3) Writing the Error Check List: Once the work piece has been selected an error check list should be constructed. The check list should contain a list of errors that could be made by the subjects while performing the task. If the chosen task is one that does not readily lend itself to assessing errors in performance then a new task should be selected. The list should be arranged in the sequence in which the errors are likely to be made. In some cases the workpiece or final product of the trainability test is also assessed for errors (e.g. Smith, 1972; Smith & Downs, 1975) although this is generally de-emphasised. A check list has the advantage that it helps the assessor to concentrate on important aspects of performance during the test period and is a quick and convenient means for recording errors.

4) Designing the Rating Scale: The range of ratings which are to be used should be decided upon and guidelines for the instructors use of the ratings developed. Most trainability tests to date have utilised a five point rating scale.

5) Designing the Instructors Script: This should inform the instructor about what to do and what materials will be needed to carry out the test. It should also provide an explanation for the applicant about the purpose of the test and include instructions to be read out while demonstrating the task.

Once each of the above steps has been satisfactorily completed the test can be administered and the individual assessed on how well he or she can perform the task that was taught.

The Typists' Trainability Assessment.

The test was designed to assess a candidate's typewriting skills and knowledge and the ability to recall and follow instructions. The subjects were required to arrange the typewriter appropriately and to adopt the correct posture. They were also required to utilise correct touch typing techniques, to vertically centre, to proofread copy and finished material and to interpret common correction marks.

The correct procedure was demonstrated by each tutor to a maximum of five subjects at once. Subjects were told that they could ask questions and would be given the opportunity to practice the various skills that had been demonstrated. After the material had been presented to the subjects they were asked to type a small passage utilising only those techniques in which they had been instructed. Subjects were informed that they could take as much time as necessary to complete the task but that it should not take any longer than half an hour. They were also told that instructors were not allowed to offer any help during this time. All subjects finished the task within the half hour guideline period. During that time they were carefully observed by tutors and any errors in performance were recorded on the error checklist. After completion of the task the number of errors was totalled and each subject received an overall rating from A (exceptional) to E (unsuitable). At this stage any other errors the subjects committed were also recorded (for a copy of the trainability test instructions and error checklist see appendices E and F).

The Work Sample.

The work sample consisted of the same passage used for the trainability test. As with the trainability test subjects were told that they were participating in a study examining methods for predicting success in training. They were then told that they would have to type a small passage which needed to be vertically centreed and required some corrections. They were instructed to be as fast and as accurate as possible and were informed that they had only 20 minutes in which to complete the task. As each subject completed the task the finishing time to the nearest minute was recorded and an identifying number known only to the experimenter placed upon the papers. The papers were then randomly allocated to each of the tutors in order to be graded. Each passage was marked out of a total of 25 with each mistake resulting in the deduction of half a mark and failure to finish the task resulting in a one mark deduction per line that was not completed (for a copy of the work sample instructions and the passage to be typed see appendices G and H).

THE META-ANALYSIS

The narrative review which is the traditional means for integrating the results from a large number of studies focusing on a particular issue has been shown to have some serious flaws (e.g. see Hunter, Schmidt, & Jackson, 1982). It is difficult for any reviewer to analyse objectively, without the aid of quantitative procedures, the results of studies drawn from large data bases. In many cases reviewers elect to analyse a sub-sample of studies (usually selected on the basis of methodological soundness and/or magnitude of results) simply to reduce the huge volume of data they must interpret. As Hunter et al (1982) point out such an approach entails the loss of a considerable amount of relevant information. Moreover, all too often, despite the smaller number of studies surveyed, the reviewer is still placed in the unenviable position of having to account for conflicting results. The reviewer may then feel obliged to provide some rationale to account for the variability between studies. Usually the presence of moderator variables is posited and the data is classified into sub-groups and re-examined. Such an approach is open to capitalization on the existence of chance relationships between study outcomes and characteristics. Hunter et al (1982) argue that the narrative review and other simple meta-analytic procedures (e.g. counting of positive and negative results, counting of significance tests, computation of means and variances across studies) fail to consider the presence of artifactual sources of variance in the data. These artifacts include sampling error, differences between studies in test reliability, differences between studies in criterion reliability and differences between studies in range restriction. Using Monte Carlo runs, they have clearly demonstrated how such artifacts obfuscate the true relationship between predictor and criterion and lead reviewers to wrong conclusions (Hunter & Hunter, 1984; Hunter et al, 1982). Using sophisticated quantitative procedures Hunter and his associates (Pearlman, Schmidt, & Hunter, 1980; Schmidt, Gast-Rosenberg, & Hunter, 1980; Schmidt, Hunter, & Pearlman, 1981; Schmidt, Hunter, & Caplan, 1981) were able to show that the between study variance in observed validity coefficients for employment tests in such diverse occupations as clerical work, computer programming and the petroleum industry were largely artifactual in nature. Such findings have tremendous implications for the field of personnel selection since they support the notion of validity generalization and contra-indicate the historical belief in the situational specificity of employment test validity. At present the Hunter et al (1982) validity generalization formulas can only correct for the artifacts noted above. There are at least three additional sources of variance that may be present in any data base that is analysed. These include differences between studies in the amount and kind of criterion contamination and deficiency, computational and typographical errors and differences in factor structure between tests of a given type (Pearlman et al, 1980; Hunter et al, 1982).

Compilation of Validity Distributions.

A thorough review of the literature on work samples was carried out. Various bibliographic indexes were used to track down relevant studies (e.g. Psychological Abstracts, Social Sciences Citation Index) and a computerised search of on-line data bases was also conducted (Psych-info, NTIS, Social Sciences Citation Index, Dissertation Abstracts, ABI/Inform) using the DIALOG Information Retrieval Service. In addition, the reference sections from relevant reports were also consulted, especially to locate early studies not included in the DIALOG files. The search procedure resulted in the collection 98 studies from which 453 usable validity coefficients were derived. It is important to note that the data base used for the present meta-analysis consists predominantly of published reports.

Information from the studies that was recorded included the correlation coefficient, the sample size, sample composition (sex and race), criterion measure used and its reliability, name of the predictor used and its reliability, date of the study and reported job title. Data from studies that failed to include a correlation coefficient, the sample size or used criterion measures other than job proficiency or training success were excluded from the analysis. Studies in which correlations were corrected for either reliability and/or range restriction were also omitted since such corrections are made during the analysis.

Certain decision rules pertaining to the nature of the validity data that should be recorded were established based on the study by Pearlman et al (1980). This was done to ensure that the information was recorded in a systematic and consistent manner. In cases where more than one relevant criterion measure was used (e.g. supervisors' ratings and production information) for a given sample the validity coefficient for each was included in the analysis. Similarly, for situations in which validity was established for two or more predictors of the same type (e.g. several types of job-related information test) for a given sample each such coefficient was recorded separately. Pearlman et al (1980) note that the inclusion of such non-independent validities results in undercorrection of the total variance making the final figures from the analysis a more conservative estimate than otherwise would have been the case. For those situations in which both rankings and ratings were provided an average was calculated and that was included in the meta-analysis.

For those studies that included validity coefficients for several dimensions of a particular criterion measure (e.g. supervisors ratings of quality of work and quantity of work) only the coefficient for the overall or summary measure was recorded, if it was available. For cases in which no such measure was available the validity coefficients for the various dimensions were averaged and that figure recorded for the analysis. When performance on a criteria was assessed by more than one group of individuals (e.g. ratings by supervisors and ratings by peers) then an average combined value was also calculated. The sample size for the cases outlined above was determined by multiplying the original number of subjects by the number of dimensions that were averaged. For reports that included analysis of sub-groups (sexual and racial) as well as providing an overall validity coefficient only data from the sub group analysis was included. The type of correlation coefficient used in each of the studies (e.g. tetrachoric, biserial, triserial, pearson) was not recorded since a previous analysis (Pearlman et al, 1980) suggested that combining such data in the analysis was likely to have only minimal effects and result in a slight underestimation of the total sampling error. Finally, for those studies that reported more than one follow-up only data from the first was included in the analysis.

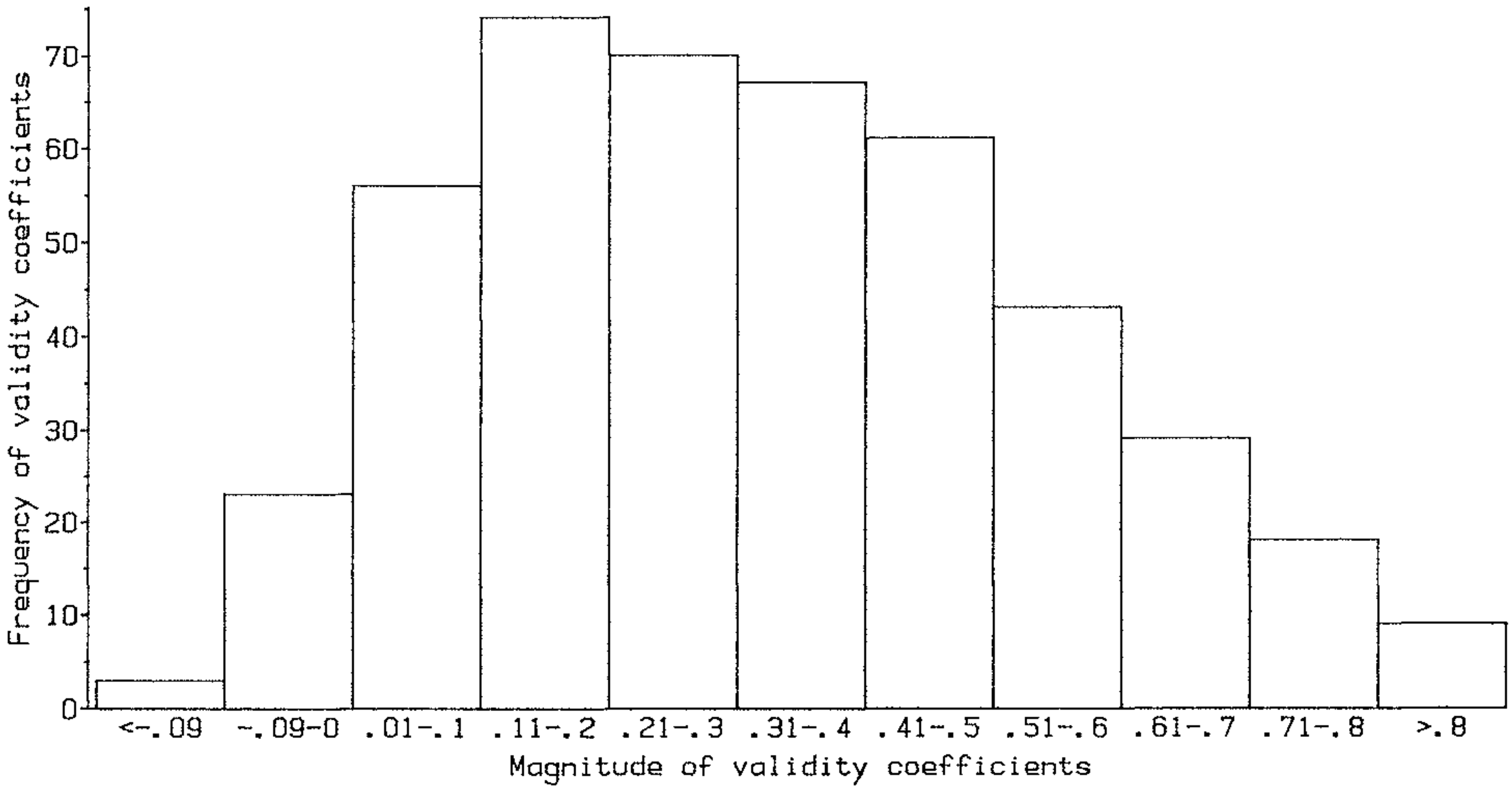

Figure 1. Distribution of validity coefficients for work samples.

Characteristics of Studies.

A total of 98 articles were located which contained information suitable for inclusion in the meta-analysis. These articles reported 204 separate correlation coefficients based on independent samples. This led to an average of approximately two recorded validity coefficients per article and a mean sample size of 90. Utilisation of the decision rules outlined above resulted in the generation of 453 usable correlation coefficients. This increased the average number of reported validity coefficients to approximately five per study and gave a mean sample size of 141. Table two includes a summary of relevant information pertaining to the characteristics of the studies and includes an analysis using the sub groups suggested by Robertson and Kandola (1982). Figure one depicts the distribution of validity coefficients used in the analysis.

TABLE 2: DESCRIPTIVE CHARACTERISTICS OF STUDIES USED IN THE META-ANALYSIS

Work Sample Category

	Train-ability tests	Psycho-motor tests	Job-Info tests	Group decision making	Individual decision making	All groups combined
Number of articles located	15	38	26	13	17	98
Number of independent correlations	27	66	68	32	51	204
Number of reported subjects	1577	5880	8091	3414	3749	18383
Range of reported sample sizes	12-102	7-633	10-965	19-969	8-516	7-969
Number of derived correlations	56	137	132	53	75	453
Total number of derived subjects	3731	14988	15264	16462	13498	63944
Average sample size for analysis	67	109	116	311	180	141
Percentage of correlations > than 0	98.2	94.2	93.2	94.3	93.3	94.3

Data Analysis.

Meta-analytic formulas developed by Hunter et al (1982) were used in the present study. Their approach entails calculating the sample size weighted mean and sample size weighted variance for the observed distribution of validity coefficients. The effect of sampling error on the distribution of validity coefficients is then estimated and subtracted from the observed variance to provide an estimate of the true variance. The population mean and variance of the distribution of true validities is obtained by multiplying the observed means and variances by a constant which corrects for criterion unreliability and for range restriction. Since data relevant to criterion reliability was only available sporadically for the individual validity studies it was necessary to use an assumed distribution of criterion reliabilities. Similarly, because the individual validity studies contained little or no information on range restriction an assumed distribution of such effects was also utilised. The assumed distributions, which are shown in tables three, four and five, are taken from Pearlman et al (1980). They note that the suggested distributions are likely to be conservative estimates and hence, will tend to result in the undercorrection of means and variances in the meta-analysis.

TABLE 3: ASSUMED DISTRIBUTION OF PROFICIENCY CRITERION RELIABILITIES FOR THE VALIDITY GENERALIZATION ANALYSES

Reliability	Relative frequency
.90	3
.85	4
.80	6
.75	8
.70	10
.65	12
.60	14
.55	12
.50	10
.45	8
.40	6
.35	4
.30	3

Following the overall analysis of available coefficients an effort was made to identify moderator variables. To that end data was grouped according to Robertson and Kandola's (1984) classification of work samples. Studies could be placed into one of five categories.

1) Psychomotor tests
2) Individual, situational decision making tests
3) Job-related information tests
4) Group discussion/decision making tests
5) Trainability tests

Sub-groups were then subjected to a meta-analysis and population means, variances and credibility intervals were calculated.

TABLE 4: ASSUMED DISTRIBUTION OF TRAINING CRITERION RELIABILITIES FOR THE VALIDITY GENERALIZATION ANALYSES

Reliability	Relative frequency
.90	15
.85	30
.80	25
.75	20
.70	4
.60	4
.50	2

TABLE 5: ASSUMED DISTRIBUTION OF RANGE RESTRICTION EFFECTS FOR THE VALIDITY GENERALIZATION ANALYSES

Prior selection ratio	SD of test	Relative frequency
1.00	10.00	5
.70	7.01	11
.60	6.49	16
.50	6.03	18
.40	5.59	18
.30	5.15	16
.20	4.68	11
.10	4.11	5

CHAPTER THREE.

RESULTS

Trainability Test/Work Sample Comparison.

Results for the comparison between the trainability test and the work sample are presented in table 6. A table of means and standard deviations for the criterion and each of the predictors is included in the appendix. Performance on both the trainability test and the work sample seems to be strongly related to the mark achieved by students in their terms test. Work sample grades and trainability test error scores appear to be particularly good predictors with the obtained correlations being highly significant. Fishers' Z transformation (Kleinbaum & Kupper, 1978) was used to test for equality of the correlation coefficients. Using a non-directional test no significant difference between any of the correlation coefficients was detected ($z = 1.86$; $p > .05$).

TABLE 6: PEARSON CORRELATIONS BETWEEN TRAINABILITY TEST/WORK SAMPLE SCORES AND THE TYPING TEST

Predictor	Correlation with grade on terms test
Trainability test ratings (N = 38)	.427~~
Trainability test errors (N = 38)	.551*
Work sample grade (N = 42)	.711*
Work sample time (N = 42)	.472~~

~~p<.005

* p<.0005

Meta-analysis.

Results from the overall meta-analysis of the work samples are presented in table 7. The work samples were categorized into two groups according to whether they incorporated proficiency criteria or training criteria. For each group the sample size weighted mean r, the sample size weighted variance, and the mean r corrected for attenuation is reported. The variance due to sampling error, the sample size weighted variance corrected for attenuation and the true variance of the corrected r's is also given. The remaining figures displayed in table 7 include the percentage of variance due to sampling error, the 90% credibility intervals and the results of a statistical test for evaluating the significance of the remaining population variance (see Hunter et al, 1982).

TABLE 7: VALIDITY GENERALIZATION RESULTS FOR PROFICIENCY AND TRAINING CRITERIA DISTRIBUTIONS

	Predictor		
	Work sample training criteria	Work sample proficiency criteria	Work sample proficiency criteria-outlier excluded
Weighted mean r	.36	.19	.27
Weighted variance of r	.0305	.0325	.0391
Mean r corrected for attenuation	.61	.40	.55
Variance corrected for attenuation	.0778	.1380	.1513
Sampling error variance	.0203	.0270	.0367
True variance of corrected r	.0575	.1110	.1146
% of variance due to sampling error	26%	20%	24%
90% credibility interval	$.22 < p < 1.00$	$-.15 < p < .95$	$-.01 < p < 1.00$
Chi square value	726.14*	1120*	824.7*

* $p < .00005$

A close examination of table 7 reveals that in all cases the estimated mean true validity following correction for attenuation was considerably greater than the average observed validity. Hunter et al (1982) suggest that mean correlations greater than two standard deviations from zero can almost always be considered positive. For the present analysis only work samples using training criteria were able to meet that requirement. That group also achieved larger mean r's (corrected and uncorrected), obtained smaller variances overall and was able to account for the largest percentage of variance due to sampling error. The work sample-proficiency criteria studies contained an outlier in terms of sample size. Following the recommendations of Hunter et al (1982) the data set was reanalysed with information from the outlier study omitted. The results suggest that the study may have unduly influenced the meta-analysis. When the study was excluded variance increased, however, there was a substantial gain in the magnitude of the mean r's and the 90% credibility interval improved. Furthermore, much of the increase in variance was offset by a corresponding increase in sampling error variance. The Chi square values displayed in table 7 for the formal test of no variation are all highly significant. Critical values were determined using a transformation suggested by Hays (1973) for Chi squares with degrees of freedom greater than 100. Hunter et al (1982) note that results obtained from the significance test should be interpreted cautiously since the test has a very high statistical power. However, in this case, the large values attained indicate that the remaining variance is non-trivial and clearly greater than that expected by chance.

Moderator Analysis.

Hunter et al (1982) have argued that if the remaining true variance following correction for statistical artifacts is large then it is indicative that moderator variables may be operating. Results from the significance tests in the overall analysis showed that a considerable amount of variance remained unaccounted for once all corrections had been implemented. This suggested that an analysis using moderator variables would be appropriate. To that end, work sample studies were classified according to Robertson and Kandola's (1982) categorization with the addition of trainability tests as a separate group. Results from the analyses are summarised in tables 8 and 9.

TABLE 8: VALIDITY GENERALIZATION RESULTS FOR MODERATOR ANALYSIS USING PROFICIENCY CRITERIA DISTRIBUTIONS

	Predictor			
	Psychomotor tests	Job-info tests	Group discussion decision making tests#	Individual decision making tests#
Weighted mean r	.22	.29	.32	.24
Weighted variance of r	.0545	.0550	.0120	.0214
Mean r corrected for attenuation	.46	.59	.64	.50
Variance corrected for attenuation	.2293	.2123	.0316	.0809
Sampling error variance	.0484	.0389	.0215	.0359
True variance of corrected r	.1809	.1734	.0101	.0450
% of variance due to sampling error	21%	18%	68%	44%
90% credibility interval	-.24<p<1.00	-.09<p<1.00	.48<p<.80	.15<p<.85
Chi square value	282.18*	381.22*	36.82(ns)	90.12~~

* p < .00005
~~ p < .005
ns Non significant
analysis conducted with outlier study omitted

TABLE 9: VALIDITY GENERALIZATION RESULTS FOR MODERATOR ANALYSIS
USING TRAINING CRITERIA DISTRIBUTIONS

	Predictor				
	Trainability tests	Psycho-motor tests	Job-info tests	Group discussion tests	Individual decision making tests
Weighted mean r	.39	.30	.42	.32	.30
Weighted variance of r	.0450	.0257	.0256	.0110	.0311
Mean r corrected for attenuation	.65	.52	.69	.56	.52
Variance corrected for attenuation	.1147	.0700	.0584	.0238	.0864
Sampling error variance	.0299	.0183	.0117	.0148	.0250
True variance of corrected r	.0848	.0517	.0467	.0090	.0614
% of variance due to sampling error	26%	26%	20%	62%	29%
90% credibility interval	.17<p<1.00	.15<p<.89	.34<p<1.00	.40<p<.72	.11<p<.93
Chi square value	176.45*	219.34*	257.88*	18.35(ns)	27.90~~

* p < .0005
~~ p < .005
ns Non significant

According to Hunter et al (1982) a moderator variable will manifest during the analysis in two ways: (1) mean correlations will differ between subgroups and (2) the true variance will be less for each subgroup than for the data base as a whole. Table 8 shows that in general, the analysis of separate work sample categories employing proficiency criteria has not resulted in any substantial differences in outcomes from the overall analysis. The major exception is the group discussion/decision making test category. This subgroup recorded a markedly larger true mean r and a reduction in the value of the true variance from that of the overall analysis. The variance that remained after correction in the meta-analysis was found to be non-significant. The only other improvement indicated in table 8 is for the group discussion and individual decision making test categories. The results show that for both these subgroups the percentage of variance accounted for by sampling error is noticeably greater and they now have 90% credibility intervals that do not include zero. For the psychomotor and job-information test subgroups the true variance actually increased. For all the subgroups, except group discussion, the remaining variance following correction for statistical artifacts is much greater than that expected by chance.

As can be seen from table 9 the meta-analysis of subgroups using training criteria shows a trend similar to that which occurred for proficiency criteria. The group discussion/decision making category was the only one to meet Hunter et al's (1982) requirements for a moderator variable. It had a lower corrected mean r than occurred for the overall analysis and the remaining true variance following corrections for statistical artifacts was found to be negligible. The 90% credibility intervals for three subgroups (Psychomotor tests, job-information tests and group discussion tests) became narrower while for the remaining two subgroups (trainability tests and individual decision making tests) they enlarged. For all subgroups, with the exception of the group discussion category, the true variance remaining after correction for statistical artifacts was highly significant.

While the meta-analysis using moderator variables may not have been particularly enlightening some useful information was generated. An examination of tables 8 and 9 reveals that the higher corrected mean r and smaller variance for the work samples using training criteria as compared to proficiency criteria from the overall analysis is carried through to the analysis by subgroups. The one exception is the group discussion category. However, for that subgroup the mean r's and true variances for training and proficiency criteria are very similar. The trend from the overall analysis for work samples using training criteria to obtain narrower credibility intervals also persists through to the meta-analysis of subgroups. The width of the 90% credibility intervals for psychomotor tests and job-information tests using training criteria was less than when performance criteria were employed. Furthermore, when using training criteria the mean for both subgroups was in excess of two standard deviations above zero which was not the case when proficiency criteria were used. For the group discussion subgroup the width of the credibility interval remained the same for both training and proficiency criteria. For the individual decision making subgroup the obtained 90% credibility interval was narrower when proficiency criteria were used.

CHAPTER FOUR.

DISCUSSION

Trainability test/work sample comparison.

Both the work sample and the trainability test were able to predict with a reasonable degree of accuracy the student's score on their end of term typing test. The grade obtained on the work sample provided the best basis for prediction, followed in turn by trainability test error scores, time taken on the work sample and trainability test ratings. While some measure of success was achieved in the use of the predictors the results did not support the hypothesis stated in the introduction. That is, trainability tests were not superior to work samples in the prediction of training success for typing students. In fact, although not significant, the trend in the data was for the work sample to achieve superior predictability compared to the trainability test. Although the obtained results were unexpected they were not inexplicable. Downs (1977) recommends that trainability tests be administered individually. She acknowledges that in some circumstances such an ideal situation will not be achievable. In such cases, she suggests that the tests should be administered to at most three applicants at a time (Downs, 1972; 1977). In the case of the present study, constraints on time and the availability of tutors meant that the trainability tests were administered to as many as five students at once. This, coupled with the tutors relative unfamiliarity with the approach, may have contributed to a loss of information and hence, to the poor predictive validity compared to the work sample. This difficulty is exacerbated by the fact that the work sample was able to be administered to larger groups and, except for grading, required little input on the part of tutors. Moreover, the tutors were able to employ their own grading system when evaluating work sample performance since the polytechnic already utilised a work sample based approach for assessments. For these reasons the grade assigned to the work sample was more likely to be an accurate and reliable indice of performance. This possibility is further substantiated when one examines closely all of the data available. Trainability test error scores were the second best predictor overall. Given the reasoning outlined above this should not be at all surprising. The number of errors committed during the trainability test can be considered a more objective and hence presumably a more reliable and valid measure than a subjective rating of the students performance. Perusal of the data from other studies that have employed trainability tests reveals that it is somewhat at odds with this explanation. One would expect that for most studies the error checklist would prove to be superior to any holistic assessments. In actual fact, global ratings are just as successful as objective error ratings in predicting training outcomes. Robertson and Downs (1979) have suggested that this apparent anomalous state of affairs may be due to deficiencies in the error checklists used. They also note that the utilisation of an error checklist ensures that the raters attention is focused on relevant behaviours which should help improve the validity of the overall rating. Returning to the present study it seems reasonable to suppose that the superiority of the work samples may be at least in part due to deficiencies in the administration and scoring of the trainability tests. However, to what extent exactly these methodological shortcomings account for discrepancies in the results remains an empirical question not readily answerable given the information available.

One further explanation that may help account for the unexpected superiority of work samples centres on the nature of the content domain the predictors were used to assess. Although a task analysis questionnaire was used as the basis for constructing the trainability test and the work sample the relevance these devices have to the final criterion differ greatly. The task analysis identified a number of behaviours which were regarded by both instructors and students as crucial elements to be learnt during the course. Examples include the acquisition of correct touch typing techniques, the adoption of appropriate posture when seated at the typewriter, the ability to proofread material etc. These important elements of the typists task were assessed to different extents by each predictor. Trainability tests, due to features of their design, were more easily able to evaluate all of these behaviours. The work samples, being subject to more constraints, were confined to assessing typing performance as measured by number of errors in copy and time taken to complete the task. The important point to recognise is that much of the information generated by the use of the trainability test was rendered superfluous. This is due to the nature of the criterion measure adopted by the polytechnic, which it can be argued bears little relationship to the assessment of many of the important behaviours identified in the task analysis. The criterion measure, the typing student's final terms test, is simply a passage requiring formatting and corrections which students are expected to transcribe. Many behaviours, such as appropriate posture and correct touch typing techniques etc, will not necessarily directly enable the students to achieve a superior grade in the terms test. This is the case because students who do not adopt these behaviours may, at least in the short term, still be able to type quickly and with few mistakes. Thus, the trainability test focussed on behaviours considered important to be instilled in students during a typing training course, unfortunately some of these behaviours were not related to performance on the criterion. Hence, much of the information gathered by the trainability test was irrelevant. On the other hand the work sample closely resembled the criterion measure used. If the criterion was altered so that it assessed the acquisition of those behaviours judged relevant to successful typing performance, or the trainability test was modified so that it measured behaviours more closely related to those assessed by the current criterion, then the results from the present study could well have been more in line with the hypothesis stated in the introduction.

The Meta-analysis.

The results from the meta-analysis are instructive when compared to Robertson and Kandola's (1982) findings. Median validity coefficients reported by Robertson and Kandola (1982) are generally markedly lower than the corrected mean validity coefficients derived from the meta-analysis, particularly for those studies using training criteria. The meta-analysis however, failed to corroborate Robertson and

Kandola's (1982) categorization of work samples. The only exception to this was the group discussion/decision making category. All other groups fell short of Hunter et al's (1982) requirements of a moderator variable. The considerable amount of variance remaining following corrections for attenuation and sampling error suggest that work samples are to some degree situationally specific. What remains to be determined is exactly what moderator variables are responsible for the variability in the data.

Despite the fact that Robertson and Kandola's (1982) distinctions between work samples were not supported empirically other useful information regarding the validity of work samples was gained. That which is most pertinent relates to the credibility intervals. All categories of work sample that employed training criteria and two of the categories which utilised proficiency criteria (group discussion/decision making; individual decision making) obtained 90% credibility intervals which did not include zero. Hunter et al (1982) have suggested that under such circumstances it is safe to assume that the correlation between predictor and criterion is always likely to be positive. In other words, despite the situational specificity of work samples their predictive validity is sufficiently robust so that in most situations they will increase the probability of accurate or correct decisions being made.

An interesting facet of the data from the meta-analysis is that unlike many earlier studies which employed this methodology (see Pearlman et al, 1980; Schmidt et al, 1980; Hunter et al, 1982) the percentage of variability in the results accounted for by sampling error was not great. Such findings are unusual in that it is more typical for the diversity in results between studies to be found to be due to the statistical artifacts enumerated by Hunter et al (1982). However, some more recent meta-analyses have also reported outcomes similar to those obtained in the present study (e.g. Linn, Harnisch, & Dunbar, 1981; Mabe & West, 1982; Steel & Ovalle, 1984). That is, they have found that the conflicting results reported in the literature are real and not entirely artifactual in nature. Subsequent analysis in these other studies has been in the most part successful in identifying moderator effects.

Demographic features of the data used in the present study may be partially responsible for some of the unanticipated findings that resulted. For example, the average number of correlations reported per study was approximately five. This means that there is some lack of independence in the reported data which in turn is likely to lead to underestimation of sampling error. Furthermore, the average sample size for the analysis was 141 and for the subgroups ranged from 67 to as many as 311. The effect of such relatively large mean N's per correlation is to also minimise the sampling error calculated in the meta-analysis. Also of some influence is the fact that only three of the seven artifacts identified by Hunter et al (1982) were taken into account during the analysis. Hence, some of the remaining variance could be further reduced if the influence of these other artifacts could be estimated and then corrected.

Despite the probable influence of these various factors on the results of the meta-analysis a realistic appraisal of the data would suggest that some moderator effect is at work. The variance remaining following the analysis is simply too great to be reasonably accounted for by the factors outlined above. Further analysis should be conducted in order to identify these moderator effects. While Robertson and Kandola's (1982) division of work samples has, for the most part, not been validated, it seems inherently plausible that variability in the data may be due to the operation of different types of work sample. The present meta-analysis lends some support to such a notion, in that, studies placed within the group discussion/decision making category were homogeneous, that is, contained very little variability in data that was not due to statistical artifacts. What is required is the development of a classificatory system that can be used to distinguish between work samples along relevant dimensions such as verbal-motor requirements, degree of relatedness to actual job activities etc. Such a system can be used on an a priori basis to assign work samples to different groupings. Meta-analytic techniques can then be used to corroborate the appropriateness of such categorizations. An alternative, and slightly more sophisticated approach, is to use multiple regression to assess the extent to which such situational factors can account for the variation across studies (e.g. Mabe & West, 1982; Steel & Ovalle, 1983).

Future Directions.

The results from the present study point to the need for more work to be carried out on trainability tests. In particular, information is required on the validity of such tests relative to other potential predictors that could be used. Such studies will provide a basis for those in personnel selection to make judgements regarding the appropriateness or otherwise, of employing a trainability test in a specific situation. Unfortunately most of the work conducted by Downs and her colleagues to date, has investigated the validity of trainability tests in isolation.

Research investigating the theory behind trainability tests is also required. If trainability tests are to be distinguished from psychomotor work samples then the theoretical basis for doing so needs to be made explicit. It is possible that a trainability test may be nothing more than a psychomotor work sample which has benefitted from some superficial cosmetic changes and a new name. On the other hand, trainability tests may in fact measure something distinct from psychomotor work samples. How and why trainability tests are able to do that are important questions that need to be answered. It is likely that the notion of point-to-point correspondence (Asher & Sciarrino, 1974), the principle that underlies psychomotor work samples, also underpins trainability tests. What needs to be established clearly is how the principle operates in the case of trainability tests. A similar plea by Robertson and Mindel (1980) has also argued for research along such lines. They believe the result of such work will be to generate important information regarding the range of jobs or training courses for which trainability tests can be applied. One promising research avenue that could shed some light on these issues centres on the work of Stern and Gordon (1961). They administered the Oral Direction Test (Langmuir, 1952; cited in Stern & Gordon, 1961) which provided a measure of an individuals ability to understand instructions. The score the subject obtained on the test was found to be highly predictive of subsequent aptitude in training. It is possible that the success of trainability tests stems from the fact that they may simply be measuring trainees ability to follow instructions.

Other research on trainability tests should also be pursued. Robertson and Mindel (1980) have suggested that one important question that should be resolved in the near future is the extent to which previous experience on the part of the applicant interferes with the value of trainability testing. No efforts have been made so far to investigate this issue. The reliability of trainability tests has been highlighted as a problem area in several reports (Robertson & Downs, 1979; Robertson & Mindel, 1980) and still remains such. The validity of overall assessments as compared to the error checklist is also a topic that merits attention. Robertson and Downs (1980) suggested that one useful area of further research connected to this would be an evaluation of the possibility of producing a statistically weighted error checklist.

As can be seen, there exists considerable scope for further research and development of trainability tests. Our knowledge of these predictors is not so great that we can afford to be complacent.

REFERENCE NOTES

1. DOWNS, S. (1982).
Work sampling techniques. In C. J. Brotherton (Ed) Towards Fairness in Selection and Placement, Wiley, London, Draft Copy.

2. DOWNS, S. (1984).
Trainability testing, economic and social benefits. Paper presented at the New Zealand Psychological Society Annual Conference, Palmerston North.

REFERENCES

ARVEY, R. D. (1979).
Unfair discrimination in the employment interview: Legal and psychological aspects. Psychological Bulletin, 86, 736-765.

ARVEY, R. D., & CAMPION, J. E. (1982).
The employment interview: A summary and review of recent research. Personnel Psychology, 35, 281-322.

ASHER, J. J., & SCIARRINO, J. A. (1974).
Realistic work sample tests: A review. Personnel Psychology, 27, 519-533.

BLOOD, M. R. (1974).
Job samples: A better approach to selection testing? American Psychologist, 29, 218-219.

BRUGNOLI, G. A., CAMPION, J. E., & BASEN, J. A. (1979).
Racial bias in the use of work samples for personnel selection. Journal of Applied Psychology, 64, 119-123.

CAMPION, J. E. (1972).
Work sampling for personnel selection. Journal of Applied Psychology, 56, 40-44.

CASCIO, W. F., & PHILLIPS, N. F. (1979).
Performance testing: A rose among thorns? Personnel Psychology, 32, 751-766.

CHADWICK-JONES, J. K., NICHOLSON, N., & BROWN, C. (1982).
Social Psychology of Absenteeism. Praeger, New York.

COHEN, S. L., & PENNER, L. A. (1976).
The rigors of predictive validation: Some comments on "A job learning approach to performance prediction." Personnel Psychology, 29, 595-600.

CRONBACH, L. J. (1966).
Essentials of psychological testing. Harper and Row, N.Y.

DOWNS, S. (1970).
Predicting training potential. Personnel Management, 2, 26-28.

DOWNS, S. (1972).
Trainability assessments: Fork truck operators. Research Paper SL4, Industrial Training Research Unit, Cambridge, England.

DOWNS, S. (1973).
Trainability assessments: Sewing machinists. Research Paper SL6, Industrial Training Research Unit, Cambridge, England.

DOWNS, S. (1977).
Trainability testing: A practical guide to selection.London; Training Services Agency.

Downs, S. (1984).
Testing Trainability. Nfer-Nelson, England.

DOWNS, S., FARR, R. M., & COLBECK, L. (1978).
Self-appraisal: A convergence of selection and guidance. Journal of Occupational Psychology, 51, 271-278.

EINHORN, H. J., & BASS, A. R. (1971).
Methodological considerations relevant to discrimination in employment testing. Psychological Bulletin, 75, 261-269.

FARR, J. L., O'LEARY, B. S., & BARTLETT, C. J. (1973). Effect of work sample test upon self-selection and turnover of job applicants. Journal of Applied Psychology, 58, 283-285.

FIELD, H. S., BAYLEY, G. A., & BAYLEY, S. M. (1977). Employment test validation for minority and non-minority production workers. Personnel Psychology, 30, 37-46.

GHISELLI, E. E. (1973). The validity of aptitude tests in personnel selection. Personnel Psychology, 26, 461-477.

GILL, R. W. T. (1979). The in-tray (in-basket) exercise as a measure of management potential. Journal of Occupational Psychology, 52, 185-197.

GOLDSTEIN, I. L. (1971). The application blank: How honest are the responses? Journal of Applied Psychology, 55, 491-492.

GORDON, M. E., & KLEIMAN, L. S. (1976). The prediction of trainability using a work sample test and an aptitude test: A direct comparison. Personnel Psychology, 29, 243-253.

GRANT, D. L. (1980). Issues in personnel selection. Professional Psychology, 11, 369-384.

GRANT, D. L., & BRAY, D. W. (1970).
Validation of employment tests for telephone company installation and repair occupations. Journal of Applied Psychology, 54, 7-14.

GUION, R. M. (1965).
Personnel testing. McGraw-Hill, N.Y.

HAKEL, M. D. (1986).
Personnel selection and placement. Annual Review of Psychology, 37, 351-380.

HAMNER, W. C., KIM, J. S., BAIRD, L., & BIGONESS, W. J. (1974).
Race and sex as determinants of ratings by potential employers in a simulated work-sampling task. Journal of Applied Psychology, 59, 705-711.

HAYS, W. L. (1973).
Statistics for the social sciences(2nd ed). Holt, Rinehart, & Winston. N.Y.

HOWARD, A. (1983).
Work samples and simulations in competency evaluation. Professional Psychology: Research and Practice, 14, 780-796.

HUNTER, J. E., & HUNTER, R. F. (1984).
Validity and utility of alternative predictors of job performance. Psychological Bulletin, 96, 72-98.

HUNTER, J. E., SCHMIDT, F. L., & JACKSON, G. B. (1982). Meta-analysis: Cumulating research findings across studies. Sage Publications, Beverly Hills.

KARREN, R. (1980). The Work Sampling Approach to Personnel Selection. U.S. Office of Personnel Management, Personnel Research Development Center, Washington, D.C.

KESSELMAN, G. A., & LOPEZ, F. E. (1979). The impact of job analysis on employment test validation for minority and non-minority accounting personnel. Personnel Psychology, 32, 91-108.

KLEINBAUM, D. G., & KUPPER, L. L. (1978). Applied regression analysis and other multivariable methods. Duxbury Press, North Scituate, Massachusetts.

LANDY, F. J. (1985). Psychology of work behavior(3rd ed). The Dorsey Press, Homewood, Illinois.

LINN, R. L., HARNISCH, D. L., & DUNBAR, S. B. (1981). Validity generalization and situational specificity: An analysis of the prediction of first year grades in law school. Applied Psychological Measurement, 5, 281-289.

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

MAYFIELD, E. C. (1964).
The selection interview: A reevaluation of published research. Personnel Psychology, 17, 239-260.

McCORMICK, E. J., & TIFFIN, J. (1975).
Industrial psychology (6th ed). George Allen and Unwin Ltd, London.

MOUNT, M. K., MUCHINSKY, P. M., & HANSER, L. M. (1977).
The predictive validity of a work sample: A laboratory study. Personnel Psychology, 30, 637-645.

MUCHINSKY, P. M. (1975).
Utility of work samples in complying with EEOC guidelines. Personnel Journal, 54, 218-220.

MUCHINSKY, P. M. (1979).
The use of reference reports in personnel selection: A review and evaluation. Journal of Occupational Psychology, 52, 287-297.

PEARLMAN, K., SCHMIDT, F. L., & HUNTER, J. E. (1980).
Validity generalization results for tests used to predict training success and job proficiency in clerical occupations. Journal of Applied Psychology, 65, 373-406.

REILLY, R. R., BROWN, B., BLOOD, M. R., & MALATESTA, C. Z. (1981). The effects of realistic previews: A study and discussion of the literature. Personnel Psychology, 34, 823-834.

REILLY, R. R., & CHAO, G. T. (1982). Validity and fairness of some alternative employee selection procedures. Personnel Psychology, 35, 1-62.

ROBERTSON, I., & DOWNS, S. (1979). Learning and the prediction of performance: Development of trainability testing in the United Kingdom. Journal of Applied Psychology, 64, 42-50.

ROBERTSON, I. T., & KANDOLA, R. S. (1982). Work sample tests: Validity, adverse impact and applicant reaction. Journal of Occupational Psychology, 55, 171-183.

ROBERTSON, I. T., & MINDEL, R. M. (1980). A study of trainability testing. Journal of Occupational Psychology, 53, 131-138.

SCHMIDT, F. L., GAST-ROSENBERG, I., & HUNTER, J. E. (1980). Validity generalization results for computer programmers. Journal of Applied Psychology, 65, 643-661.

SCHMIDT, F. L., GREENTHAL, A. L., HUNTER, J. E., BERNER, J. G., & SEATON, F. W. (1977).
Job-sample vs paper-and-pencil trades and technical tests: Adverse impact and examinee attitudes. Personnel Psychology, 30, 187-197.

SCHMIDT, F. L., & HUNTER, J. E. (1981).
Employment testing: Old theories and new research findings. American Psychologist, 36, 1128-1137.

SCHMIDT, F. L., HUNTER, J. E., & CAPLAN, J. R. (1981).
Validity generalization results for job groups in the petroleum industry. Journal of Applied Psychology, 66, 261-273.

SCHMIDT, F. L., HUNTER, J. E., McKENZIE, R. C., & MULDROW, T. W. (1979).
Impact of valid selection procedures on work-force productivity. Journal of Applied Psychology, 64, 609-626.

SCHMIDT, F. L., HUNTER, J. E., & PEARLMAN, K. (1981).
Task differences as moderators of aptitude test validity in selection: A red herring. Journal of Applied Psychology, 66, 166-185.

SCHMITT, N. (1976).
Social and situational determinants of interview decisions: Implications for the employment interview. Personnel Psychology, 29, 79-101.

SIEGEL, A. I. (1983).
The miniature job training and evaluation approach: Additional findings. Personnel Psychology, 36, 41-57.

SIEGEL, A. I., & BERGMAN, B. A. (1975).
A job learning approach to performance prediction. Personnel Psychology, 28, 325-339.

SMITH, M. C. (1972).
Trainability assessments: Electronic assemblers. Research Paper SL5, Industrial Research Training Unit, Cambridge, England.

SMITH, M. C. (1977).
A comparison of the value of trainability assessments and other tests for predicting the practical performance of dental students. International Review of Applied Psychology, 25, 125-130.

SMITH, M. C., & DOWNS, S. (1975).
Trainability assessments for apprentice selection in shipbuilding. Journal of Occupational Psychology, 48, 39-43.

STEEL, R. P., & OVALLE, N. K. (1984).
A review and meta-analysis of research on the relationship between behavioral intentions and employee turnover. Journal of Applied Psychology, 69, 673-686.

STERN, F., & GORDON, L. V. (1961).
Ability to follow instructions as a predictor of success in recruit training. Journal of Applied Psychology, 45, 22-24.

TENOPYR, M. L., & OELTJEN, P. D. (1982).
Personnel selection and classification. Annual Review of Psychology, 33, 581-618.

THORNTON, G. C., & BYHAM, W. C. (1982).
Assessment centers and managerial performance. Academic Press, N.Y.

WAGNER, R. (1949).
The employment interview: A critical summary. Personnel Psychology, 2, 17-46.

WANOUS, J. P. (1977).
Organizational entry: Newcomers moving from outside to inside. Psychological Bulletin, 84, 601-618.

WERNIMONT, P. F. (1962).
Re-evaluation of a weighted application blank for office personnel. Journal of Applied Psychology, 46, 417-419.

WERNIMONT, P. F., & CAMPBELL, J. P. (1968).
Signs,samples and criteria. Journal of Applied Psychology, 52, 372-376.

WRIGHT, O. R. (1969).

Summary of research on the selection interview since 1964. Personnel Psychology, 22, 341-413.

APPENDIX A

SOURCES OF VALIDITY DATA FOR THE META-ANALYSIS

ANDERSON, A. V., & FRIEDMAN, S. (1952).
Prediction of performance in a navy dental prosthetic technician training course. American Psychologist, 7, 288.

ANONYMOUS. (1954).
Validity information exchange. No 7-094. Personnel Psychology, 7, 572.

ARBOUS, A. G., & MAREE, J. L. (1951).
Contribution of 2-group discussion techniques to a validated test battery. Occupational Psychology, 25, 73-89.

ASH, P. (1960).
Validity information exchange. No 13-07. Personnel Psychology, 13, 456.

BAIER, D. E., & DUGAN, R. D. (1956).
Tests and performance in a sales organisation. Personnel Psychology, 9, 17-26.

BASS, B. M., & COATES, C. H. (1954).
Validity information exchange. No 7-082. Personnel Psychology, 7, 553-554.

BENDER, W. R. G., & LOVELESS, H. E. (1958).
Validation studies involving successive classes of trainee stenographers. Personnel Psychology, 11, 491-508.

BENNETT, G. K., & FEAR, R. A. (1943).
Mechanical comprehension and dexterity. Personnel Journal, 22, 12-17.

BLUM, M. L. (1940).
A contribution to manual aptitude measurement in industry: The value of certain dexterity measures for the selection of workers in a watch factory. Journal of Applied Psychology, 24, 381-416.

BLUM, M. L. (1943).
Selection of sewing machine operators. Journal of Applied Psychology, 27, 35-40.

BRASS, D. J., & OLDHAM, G. R. (1976).
Validating an in-basket test using an alternative set of leadership scoring dimensions. Journal of Applied Psychology, 61, 652-657.

BRAY, D. W., & GRANT, D. L. (1966).
The assessment center in the measurement of potential for business management. Psychological Monographs, 80, Whole No. 625.

BRIDGEMAN, C. S., SPAETHE, M., & DIGNAN, F. (1958).
Validity information exchange. No 11-21. Personnel Psychology, 11, 264-265.

CAMPION, J. E. (1972).
Work sampling for personnel selection. Journal of Applied Psychology, 56, 40-44.

CANDEE, B., & BLUM, M. L. (1937).
Report on a study done in a watch factory. Journal of Applied Psychology, 21, 572-582.

CARPENTER, C. R., GREENHILL, L. P., HITTINGER, W. F., McCOY, E. P., McINTYRE, C. J., MURNIN, J. A., & WATKINS, R. W. (1954).
The development of a sound motion proficiency test. Personnel Psychology, 7, 509-523.

CASTLE, P. F. C., & GARFORTH, F. I. (1951).
Selection, training and status of supervisors: 1 Selection. Occupational Psychology, 25, 108-124.

CROFT, E. J. (1959).
Prediction of clothing construction achievement of high school girls. Educational and Psychological Measurement, 19, 653-655.

DOWNS, S. (1970).
Predicting training potential. Personnel Management, 2, 26-28.

DOWNS, S. (1972).
Trainability assessments: Fork truck operators. Research Paper SL4, Industrial Training Research Unit, Cambridge, England.

DOWNS, S. (1973).
Trainability assessments: Sewing machinists. Research Paper SL6, Industrial Training Research Unit, Cambridge, England.

DOWNS, S. (1982).
Work sampling techniques. In C. J. Brotherton (Ed) Towards Fairness in Selection and Placement, London, Wiley, Draft Copy.

DUBOIS, P. H., & WATSON, R. I. (1950).
The selection of patrolmen. Journal of Applied Psychology, 34, 90-95.

DURRETT, H. L. (1961).
Validity information exchange. No 14-03. Personnel Psychology, 14, 453-455.

FLEISHMAN, E. A. (1955).
Predicting code proficiency of radiotelegraphers by means of aural tests. Journal of Applied Psychology, 39, 150-155.

FOREHAND, G. A., & GUETZKOW, H. (1961).
The administrative judgement test as related to descriptions of executive judgement behaviour. Journal of Applied Psychology, 45, 257-261.

FREDERIKSEN, N., SAUNDERS, D. R., & WAND, B. (1957).
The in-basket test. Psychological Monographs: General and Applied, 71, Whole No. 438.

GAEL, S., & GRANT, D. L. (1972).
Employment test validation for minority and non-minority telephone company service representatives. Journal of Applied Psychology, 56, 135-139.

GAEL, S., GRANT, D. L., & RITCHIE, R. J. (1975).
Employment test validation for minority and non-minority telephone operators. Journal of Applied Psychology, 60, 411-419.

GAEL, S., GRANT, D. L., & RITCHIE, R. J. (1975).
Employment test validation for minority and non-minority clerks with work sample criteria. Journal of Applied Psychology, 60, 420-426.

GIESE, W. J. (1949).
A tested method for the selection of office personnel. Personnel Psychology, 2, 525-545.

GLANZ, E. (1949).
A trade test for power sewing machine operators. Journal of Applied Psychology, 33, 436-441.

GLASER, R., SCHWARZ, P. A., & FLANAGAN, J. C. (1958).
The contribution of interview and situational performance procedures in the selection of supervisory personnel. Journal of Applied Psychology, 42, 69-73.

GLEASON, W. J. (1957).
Predicting army leadership ability by modified leaderless group discussion. Journal of Applied Psychology, 41, 231-235.

GLICKMAN, A. S. (1956).
The naval knowledge test. Journal of Applied Psychology, 40, 389-392.

GLICKMAN, A. S., FARR, J. L., O'LEARY, B. S., PFEIFFER, C. M., GOLDSTEIN, I. L., & BARTLETT, C. J. (1971).
Ethnic group membership as a moderator in the prediction of job performance: An examination of some less traditional predictors. Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, Department of the Navy, Washington D. C.

GOEBEL, R. A., BAUM, D. R., & HAGIN, W. V. (1971).
Using a ground trainer in a job sample approach to predicting pilot performance. Air Force Human Resources Laboratory, Flying Training Division, Williams Air Force Base, Arizona.

GORDON, L. (1967).
Clinical, psychometric and work sample approaches in the prediction of success in peace corps training. Journal of Applied Psychology, 51, 111-119.

GORDON, M. E., & COHEN, S. L. (1973).
Training behaviour as a predictor of trainability. Personnel Psychology, 26, 261-272.

GORDON, M. E., & KLEIMAN, L. S. (1976).
The prediction of trainability using a work sample test and an aptitude test: A direct comparison. Personnel Psychology, 29, 243-253.

GRIGG, A. E. (1948).
A farm knowledge test. Journal of Applied Psychology, 32, 452-455.

HAKSTIAN, R. A., WOOLSEY, L. K., & SCHROEDER, M. L. (1986).
Development and application of a quickly scored in-basket exercise in an organisational setting. Educational and Psychological Measurement, 46, 385-396.

HANDYSIDE, J. D., & DUNCAN, D. C. (1954).
Four years later: A follow-up of an experiment in selecting supervisors. Journal of Occupational Psychology, 28, 2-23.

HARRIS, D. H. (1964).
Development and validation of an aptitude test for inspectors of electronic equipment. Journal of Industrial Psychology, 2, 29-35.

HEARNSHAW, L. S. (1937).
Selection tests for paper sorters. Human Factor, London, 11, 146-153.

HOLLENBECK, G. P., & MCNAMARA, W. J. (1965).
Cucpat and programming aptitude. Personnel Psychology, 18, 101-106.

INSKEEP, G. C. (1971).
The use of psychomotor tests to select sewing machine operators - some negative findings. Personnel Psychology, 24, 707-714.

KESSELMAN, G. A., & LOPEZ, F. E. (1979).
The impact of job analysis on employment test validation for minority and non-minority accounting personnel. Personnel Psychology, 32, 91-108.

KNAUFT, E. B. (1954).
Validity information exchange. No 7-070. Personnel Psychology, 7, 405-406.

KRIEDT, P. H. (1952).
Validation of a correspondence aptitude test. Journal of Applied Psychology, 36, 5-7.

LAUER, A. R. (1955).
Comparison of group paper and pencil tests with certain psychological tests for measuring driving aptitude of army personnel. Journal of Applied Psychology, 39, 318-321.

LAYTON, W. L. (1953).
Predicting success in dental school. Journal of Applied Psychology, 37, 251-255.

LONG, W. F., & LAWSHE, C. H. (1947).
The effective use of manipulative tests in industry. Psychological Bulletin, 44, 130-148.

MANDELL, M. M. (1947).
The selection of foremen. Educational and Psychological Measurement, 7, 385-397.

MANDELL, M. M. (1950).
Scientific selection of engineers. Personnel, 26, 296-298.

MANDELL, M. (1950).
Selecting chemists for the federal government. Personnel Psychology, 3, 53-56.

MANDELL, M. (1950).
The administrative judgement test. Journal of Applied Psychology, 34, 145-147.

MANDELL, M. (1950).
Validation of group oral performance test. *Personnel Psychology*, *3*, 179-185.

MANDELL, M. M. (1956).
Validity information exchange. No 9-2. *Personnel Psychology*, *9*, 105.

McNAMARA, W. J., HUGHES, J. L. (1961).
A review of research on the selection of computer programmers. *Personnel Psychology*, *14*, 39-51.

MEYER, H. M. (1970).
The validity of the in-basket test as a measure of managerial performance. *Personnel Psychology*, *23*, 297-307.

MOORE, B. G. R., & PEEL, E. A. (1951).
Predicting aptitude for dentistry. *Occupational Psychology*, *25*, 192-199.

MOUNT, M. K., MUCHINSKY, P. M., & HANSER, L. M. (1977).
The predictive validity of a work sample: A laboratory study. *Personnel Psychology*, *30*, 637-645.

NAYLOR, G. F. K. (1954).
Aptitude tests for airtraffic control officers. *Occupational Psychology*, *28*, 209-217.

PERSONNEL RESEARCH BRANCH, THE ADJUTANT GENERAL'S OFFICE, DEPARTMENT OF THE ARMY. (1957).
Validity information exchange. No 10-41. _Personnel Psychology_, _10_, 485-486.

PERSONNEL RESEARCH BRANCH, THE ADJUTANT GENERAL'S OFFICE, DEPARTMENT OF THE ARMY. (1957).
Validity information exchange. No 10-42. _Personnel Psychology_, _10_, 487-489.

PERSONNEL RESEARCH BRANCH, THE ADJUTANT GENERAL'S OFFICE, DEPARTMENT OF THE ARMY. (1958).
Validity information exchange. No 11-16. _Personnel Psychology_, _11_, 251-253.

PERSONNEL RESEARCH BRANCH, THE ADJUTANT GENERAL'S OFFICE, DEPARTMENT OF THE ARMY. (1958)
Validity information exchange. No 11-18. _Personnel Psychology_, _11_, 257-259.

PORUBEN, A. (1950).
A test battery for actuarial clerks. _Journal of Applied Psychology_, _34_, 159-162.

ROBERTSON, I. T., & MINDEL, R. M. (1980).
A study of trainability testing. _Journal of Occupational Psychology_, _53_, 131-138.

RUBENOWITZ, S. (1958).
Predicting academic success: A follow-up study. Occupational Psychology, 32, 162-170.

SAUNDERS, W. J., SEIL, W. R., & ROSENSTEEL, R. K. (1956).
Validity information exchange. No. 9-34. Personnel Psychology, 9, 379.

SHULTZ, M. (1955).
Validity information exchange. No. 8-09. Personnel Psychology, 8, 118-119.

SKULA, M., & SPILLANE, R. F. (1954).
Validity information exchange. No 7-016. Personnel Psychology, 7, 147-148.

SMITH, K. R. (1940).
Work sample, dexterity tests, and ratings in the prediction of a motor task performance. Psychological Bulletin, 37, 448.

SMITH, M. C. (1972).
Trainability assessments: Electronic assemblers. Research Paper SL5, Industrial Research Training Unit, Cambridge, England.

SMITH, M. C. (1977).
A comparison of the value of trainability assessments and other tests for predicting the practical performance of dental students.
International Review of Applied Psychology, 25, 125-130.

SMITH, M. C., & DOWNS, S. (1975).
Trainability assessments for apprentice selection in shipbuilding.
Journal of Occupational Psychology, 48, 39-43.

SNYDER, R. (1955).
Validity information exchange. No 8-14. Personnel Psychology, 8, 263.

SPEER, G. S. (1957).
Validity information exchange. No 10-5. Personnel Psychology, 10, 80.

STEEL, M., BALINSKY, B., & LANG, H. (1945).
A study on the use of a work sample. Journal of Applied Psychology, 29, 14-21.

THORNTON, G. C., & BYHAM, W. C. (1982).
Assessment centers and managerial performance. Academic Press, N.Y.

TIFFIN, J., & GREENLY, R. J. (1939).
Experiments in the operation of a punch press. Journal of Applied Psychology, 23, 450-460.

TURNAGE, J. J., & MUCHINSKY, P. M. (1984).
A comparison of the predictive validity of assessment center evaluations versus traditional measures in forecasting supervisory job performance: Interpretive implications of criterion distortion for the assessment paradigm. Journal of Applied Psychology, 69, 595-602.

UHLMANN, F. W. (1962).
A selection test for production machine operators. Personnel Psychology, 15, 287-293.

WEBB, S. C. (1956).
The prediction of achievement for first year dental students. Educational and Psychological Measurement, 16, 543-548.

WEISLOGEL, R. L. (1954).
Development of situational tests for military personnel. Personnel Psychology, 7, 492-497.

WEISS, I. (1952).
Prediction of academic success in dental school. Journal of Applied Psychology, 36, 11-14.

WEST, L. J., & BOLANOVICH, D. J. (1963).
Evaluation of typewriting proficiency training. Journal of Applied Psychology, 47, 403-407.

WOLLOWICK, H. B., & MCNAMARA, W. J. (1969).
Relationship of the components of an assessment centre to management success. Journal of Applied Psychology, 53, 348-352.

WORBOIS, G. M. (1975).
Validation of externally developed assessment procedures for the identification of supervisory potential. Personnel Psychology, 28, Spring.

APPENDIX B

TYPISTS TASK ANALYSIS QUESTIONNAIRE

PLEASE CONSIDER THE TASKS PRESENTED ON THE FOLLOWING PAGES.

YOU SHOULD RATE THESE TASKS ACCORDING TO HOW IMPORTANT THEY ARE IN DISTINGUISHING A WELL TRAINED TYPIST FROM A POORLY TRAINED TYPIST.

EACH TASK SHOULD BE RATED ACCORDING TO THE FOLLOWING CRITERIA:

1 VERY IMPORTANT

2 IMPORTANT

3 UNIMPORTANT

4 VERY UNIMPORTANT

Thus, in the case of item one if you believe students who have done well in your course will correctly care for and maintain their typewriters, while students who have done poorly will not, then you should assign it a value of one or two.

On the other hand, if good students and poor students alike show the same degree of care and maintenance of their typewriters at the end of the course then it should be assigned a value of three or four.

If in your opinion some of the tasks are not relevant to the course you teach then you should not assign them any value.

1. Correctly cares for and maintains his/her typewriter.

2. Ensures that his/her typewriter is prepared before beginning typing.

3. Uses correct touch typing techniques.

4. Maintains correct posture when typing.

5. Arranges chair, typewriter and desk in appropriate positions.

6. Positions hands and arms correctly.

7. Uses correct keystroking technique.

8. Uses correct space bar technique.

9. Uses correct line spacing technique.

10. Uses correct backspacing technique.

11. Uses correct shift key technique.

12. Uses correct tabulating technique.

13. Can identify paper names and sizes.

14. Can vertically center.

15. Can space correctly.

16. Can proofread accurately.

17. Can realign incomplete typescripts.

18. Can type in capitals correctly.

19. Can underscore correctly.

20. Can horizontally center.

21. Can type accurately from handwritten copy.

22. Can tabulate correctly.

23. Can type on ruled lines.

24. Can understand and employ both open and closed punctuation.

25. Can locate punctuation signs on the key board.

26. Can accurately type punctuation signs and use correct spacing.

27. Is competant in typing miscellaneous characters.

28. Can select the right warm up drills and use them in a systematic manner.

29. Can correctly determine line endings.

30. Can maintain an even right hand margin.

31. Can correctly divide words at line endings.

32. Can utilise the various correction methods.

33. Is able to spread and squeeze characters.

34. Is able to interpret common correction marks used when editing copy.

35. Is able to interpret and type Roman numerals.

36. Is able to type money columns and total lines in an acceptable style.

37. Knows how to use and type metric units.

38. Can arrange the various parts of a typescript to give a clear, balanced and attractive appearance to the finished work.

39. Knows how to type the three basic styles of paragraph.

40. Knows how to type lettered or numbered paragraphs.

41. Can correctly calculate the positions of the margin and tab stops when typing open tables using both the forward space and backspace methods.

42. Can type tables with both blocked and centered column headings.

43. Can type headings of more than one line

44. Can understand the purpose of invoices and can type them correctly.

45. Can assemble carbon packs.

46. Can make corrections on carbon copies.

47. Can correctly type interoffice memorandums.

48. Can type telegrams suitable for lodging over a post office counter or by telephone.

49. Can address envelopes correctly.

50. Is familiar with the parts of a typical business letter.

51. Is knowledgable about the common letter styles and letter punctuation styles.

52. Can correctly fold letters to fit envelopes of different sizes.

53. Can type personal business letters.

54. Can type a personal information sheet.

55. Can type letters of application and can complete application forms.

56. Has a high level of production speed.

57. Is very accurate when typing.

If there are any other tasks that you believe clearly distinguish between well trained and poorly trained typists please write them in below.

I would now like you to re-examine the tasks outlined above and place them in some order of importance.

The number of the most important task should be entered first, the number of the next most important should be entered second, and so on.

You need only identify the 15 most important tasks taught in your course.

A. ____

B. ____

C. ____

D. ____

E. ____

F. ____

G. ____

H. ____

I. ____

J. ____

L. ____

M. ____

N. ____

O. ____

APPENDIX C

PROCEDURAL FLOW CHART USED BY TUTORS WHO ADMINISTERED THE WORK SAMPLE

Hand out the work sample biographical data sheet and ask students to fill it in.

↓

Hand out the passage to be typed.

↓

Read out the work sample instructions.

↓

Commence the typing task.

↓

Students have a maximum of twenty minutes to complete the task.

↓

As each student finishes collect their work and the biographical data sheet.

↓

On the bottom of the biographical data sheet note the time taken by each student to complete the work sample.

↓

Thank the students for their participation and ask them to refrain from discussing the study with other students.

↓

↓

Collect up the passage the students had to type.

APPENDIX D

PROCEDURAL FLOW CHART USED BY TUTORS WHO ADMINISTERED THE TRAINABILITY TEST

Fill in the biographical data sheet for each student (this is attached to the error check list).

↓

Read out the first section of instructions for the trainability test.

↓

Explain and demonstrate the five skills the students are expected to learn.

↓

After explaining each skill allow the students the opportunity to practice and ask questions.

↓

Read the remainder of the instructions for the trainability test.

↓

Hand out the passage to be typed.

↓

Allow the students to commence typing.

↓

Fill in the error check list and note any other errors students made.

↓

↓

Collect in the passage and the students copy and note those students who failed to complete the task within the given time limit (one hour).

↓

Thank the students for their participation and ask them to refrain from discussing the study with other students.

APPENDIX E

TRAINABILITY TEST INSTRUCTIONS

You have been asked to participate in a study looking at methods for predicting success in training.

I am going to describe to you some basic typing skills and demonstrate their use.

You will be shown how to arrange your workspace appropriately and how to sit at the typewriter.

You will also be shown some basic touch typing techniques.

I will then explain to you how to vertically centre a piece of work and will describe a few common correction marks.

During the explanation you may ask as many questions as you wish in order to be sure you understand what you have to do. You will also be given the opportunity to practice the various skills described. When I have finished explaining you will be asked to type a small passage using the techniques I have demonstrated.

When you are typing this passage I will not be able to offer any help so it is important that you pay careful attention to my explanations.

Even if you have done some typing before you will need to watch very carefully and use the methods you are shown.

1. ARRANGING THE TYPEWRITER AND POSTURE:

i) sit slightly to the right of center of the keyboard.

ii) feet should be flat on the floor, slightly apart and one foot a little in front of the other.

iii) ensure that the lower back is supported firmly with the back of the chair.

iv) sit erect with shoulders and head leaning slightly forward.

v) position the typewriter so that the front edge is in line with the edge of the desk.

vi) copy should be placed on the opposite side of the carriage return lever in a natural line of vision.

AS YOU EXPLAIN THESE POINTS YOU SHOULD DEMONSTRATE THEIR CORRECT APPLICATION. ONCE YOU HAVE FINISHED THE DEMONSTRATION THE STUDENTS SHOULD BE ALLOWED THE OPPORTUNITY TO PRACTICE.

2. TOUCH TYPING TECHNIQUES:

i) fingers should be placed lightly on the keys.

ii) the wrists should be parallel to the slope of the keyboard and the thumbs positioned close to the space bar.

iii) hands should not be resting on the framework of the typewriter.

iv) the typewriter keys should not be banged but should be struck sharply in the center using only the tips of the fingers and then released quickly.

v) the right hand should operate the right half of the keyboard and the left hand should operate the left half of the keyboard.

vi) typing should be done with the fingers - the hands and arms should be almost motionless.

vii) the letters A, S, D, F, should be used as guide keys for the left hand and the letters J, K, L, should be used as guide keys for the right hand.

viii) fingers should be returned to the guide position after striking a key on another row.

ix) the line spacer and carriage return should be operated without looking up.

AS YOU EXPLAIN THESE POINTS YOU SHOULD DEMONSTRATE THEIR CORRECT APPLICATION. ONCE YOU HAVE FINISHED THE DEMONSTRATION THE STUDENTS SHOULD BE ALLOWED THE OPPORTUNITY TO PRACTICE.

3. VERTICAL CENTERING:

i) you count the number of lines that can be fitted on a page by operating the carriage return on single spacing.

ii) count the number of lines in the material to be typed including blanks.

iii) determine how much space is left and divide by two.

iv) begin typing in the next line below the top margin.

YOU SHOULD DEMONSTRATE THIS PROCEDURE CLEARLY AND THEN ALLOW THE STUDENTS TO PRACTICE ONE OR TWO EXAMPLES.

4. PROOFREADING:

i) emphasise the necessity for proofreading both the copy material and the finished product.

5. INTERPRETING COMMON CORRECTION MARKS:

i) the following correction marks should be taught;

-alter letter(s) or word(s).

-leave space.

-transpose words.

-insert letter.

-use capital letters.

-type in full.

-move to the left.

AFTER DESCRIBING THE CORRECTION MARKS YOU SHOULD TEST THE STUDENTS BRIEFLY IN ORDER TO ENSURE THAT THEY UNDERSTAND THEIR MEANING.

Now that you are familiar with some of the basic typing skills you will be required to demonstrate these while typing a brief passage. The passage should be vertically centred and a number of corrections will have to be made.

It is important that you use only those techniques that you have been taught.

You can have as much time as necessary to complete the task however, it should not take any longer than half an hour.

During this time you may not ask any questions of the instructor.

Before we begin are there any queries?

APPENDIX F

TRAINABILITY TEST ERROR CHECKLIST

SUBJECTS

		S1	S2	S3	S4	S5
Work station/ Posture	Sits in the wrong position					
	Feet placed incorrectly					
	Lower back unsupported					
	Leans forward or leans back					
	Front edge of typewriter misaligned					
Proofreading	Does not proofread copy material					
Touch typing techniques	Fingers rest heavily on keys					
	Keys are banged					
	Slow to release keys					
	Wrists angled incorrectly					
	Thumbs positioned poorly					
	Hand resting on framework of typewriter					
	Excessive movement of hands and arms					
	Left hand used on right half of keyboard					
	Right hand used on left half of keyboard					
	Fingers not positioned on guide keys					
	Fingers not returned to guide keys					
	Looks down to operate line spacer					
	Looks down to operate carriage return					
Vertical centering	Does not count No.lines that fit on a page					
	Does not count No.lines in material to type					
	Does not divide remaining space by two					
	Does not type in next line below top margin					
Interpreting correction marks	Fails to alter letter(s)/word(s)					
	Fails to leave space					
	Fails to transpose word(s)					
	Fails to insert letters					
	Fails to change to upper case					
	Fails to type in full					
	Fails to move text to the left					
Proofreading	Does not proofread typed material					
	TOTAL ERRORS					
	CORRECTION OF ERRORS					
	OVERALL ASSESSMENT					

The subject generally notices his/her errors and subsequently corrects...

A. Always
B. Most times
C. Occaisionally
D. Never

A. Exceptional. The assessor would expect him/her to become a very good typist in a short time.
B. Above average. The assessor would expect him/her to become a good typist in a reasonable time.
C. Average. The assessor would expect him/her to become a competent typist by the end of training.
D. Below average. The assessor would expect him/her to have difficulty and to take longer to achieve a reasonable standard of typing.
E. Unsuitable. This individual would not be expected to become proficient at typing even if given considerable

Please note below any other errors the subjects committed.

Subject One:

Subject Two:

Subject Three:

Subject Four:

Subject Five:

APPENDIX G

WORK SAMPLE INSTRUCTIONS

You have been asked to participate in a study looking at methods for predicting success in training.

You will be presented with a brief passage which has to be typed out. The passage should be vertically centred and a number of corrections will have to be made. If you are unable to vertically centre or are unfamiliar with the correction marks used in the copy simply do the best that you can.

You should endeavour to be as fast and as accurate as possible.

You will have twenty minutes in which to complete the task.

APPENDIX H

PASSAGE TO BE TYPED BY SUBJECTS

Ergonomics developed through the interests of a number of different professions, and it still remains a multidisciplinary field of study. It crosses the boundries between many sceintific and professional disciplines and draws onthe data, findings and principles of each. Present-day ergonomics is an amalgam of physiology, anatomy, and medicine as one branch; physiological and experimental psychology as another; and physics and engineering as a third. The BIOLOGICAL SCIENCES provide information about the structure of the body: the operator's limitations and physical capabilities; the dimensions ofhis body; how much he can lift; the physical presures he can endure, etc. Physiological psychology deals with the functioning of the brain and the nervous system as they determine behavior; while EXPERIMENTAL PSYCHOLOGISTS attempt to understand the basic ways in which the individual uses his body to behave, to perceive, to learn, to remember, to control his motor processes, etc.

Finally, PHYSICS and ENGINEERING provide similar info about the machine and the environment with which the operator has to contend.

APPENDIX I

MEANS AND STANDARD DEVIATIONS FOR THE SUBJECTS SCORES ON THE PREDICTORS AND THE CRITERION

TABLE 10. MEANS AND STANDARD DEVIATIONS FOR THE PREDICTORS AND THE CRITERION

	Mean Score	Standard Deviation
Trainability test rating (N=38)	2.29	0.77
Trainability test error score (N=38)	5.08	4.83
Work sample error score (N=42)	16.46	5.70
Work sample-time in secs (N=42)	909.67	234.62
Work sample-terms test mark (N=42)	68.21	14.93
Trainability test-terms test mark (N=38)	68.22	21.28
Entire class-terms test mark (N=83)~	67.98	18.25

~ Includes late enrolements

APPENDIX C

PROCEDURAL FLOW CHART USED BY TUTORS WHO ADMINISTERED THE WORK SAMPLE

Hand out the work sample biographical data sheet and ask students to fill it in.

↓

Hand out the passage to be typed.

↓

Read out the work sample instructions.

↓

Commence the typing task.

↓

Students have a maximum of twenty minutes to complete the task.

↓

As each student finishes collect their work and the biographical data sheet.

↓

On the bottom of the biographical data sheet note the time taken by each student to complete the work sample.

↓

Thank the students for their participation and ask them to refrain from discussing the study with other students.

↓

↓

Collect up the passage the students had to type.

APPENDIX D

PROCEDURAL FLOW CHART USED BY TUTORS WHO ADMINISTERED THE TRAINABILITY TEST

Fill in the biographical data sheet for each student (this is attached to the error check list).

↓

Read out the first section of instructions for the trainability test.

↓

Explain and demonstrate the five skills the students are expected to learn.

↓

After explaining each skill allow the students the opportunity to practice and ask questions.

↓

Read the remainder of the instructions for the trainability test.

↓

Hand out the passage to be typed.

↓

Allow the students to commence typing.

↓

Fill in the error check list and note any other errors students made.

↓

↓

Collect in the passage and the students copy and note those students who failed to complete the task within the given time limit (one hour).

↓

Thank the students for their participation and ask them to refrain from discussing the study with other students.