VALIDITY EVIDENCE OF A MULTIPLE-CHOICE TEST AND A PERFORMANCE TEST IN AN EMPLOYMENT SETTING

by

Robert E. Lofgren

BS, University of Pittsburgh, 1992

MA, University of New Haven, 1994

MA, University of Pittsburgh, 2001

Submitted to the Graduate Faculty of

School of Education in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2005

UNIVERSITY OF PITTSBURGH

SCHOOL OF EDUCATION

This dissertation was presented

by

Robert E. Lofgren

It was defended on

November 10, 2005

and approved by

Clement A. Stone, Ph.D., School of Education

Louis A. Pingel, Ph.D., School of Education

James J. Irrgang Ph.D., School of Health and Rehabilitation Sciences

Dissertation Director: Suzanne Lane, Ph.D., School of Education

VALIDITY EVIDENCE OF A MULTIPLE-CHOICE TEST AND A PERFORMANCE TEST IN AN EMPLOYMENT SETTING

Robert E. Lofgren, Ph.D.

University of Pittsburgh, 2005

A selection procedure consisting of both multiple-choice (MC) paper-and pencil and hands on performance assessment (PA) elements was developed for a large consumer products manufacturing company in the southeastern United States for the purpose of ensuring that workers possessed the necessary knowledge, skills, and abilities required for work at a new facility. Two 125-item alternate tests using an MC format and 7 PA exercises were initially developed for the job referred to in the present study as L2/L3 Production Technician. The purpose of this study was to examine the validity evidence for the two alternate multiple-choice (MC) job knowledge tests and seven performance assessment (PA) exercises that were developed for employment selection purposes. The study sample included 432 Form A and 324 Form B examinees who took both the MC test and the PA exercises. Factor analysis results revealed that the same construct, labeled as *applied mechanical knowledge,* was measured by both the MC tests and the PA exercises. Item and test analysis results supported the use of Form A and Form B as alternate test forms. The decision consistency between the MC tests and the PA exercises did not appear to be sufficient to recommend that either form of the MC test alone could be used to select qualified L2/L3 Production Technicians. The correlations between MC score and PA total score were .627 for Form A and .612 for Form B. As part of a content analysis, subject-matter experts rated a large number of MC items as either having “no

relationship" or "small relationship" to the PA exercises. However, subject-matter experts did rate the PA exercises as having a great importance to the job of L2/L3 Production Technician.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

1.1 OVERVIEW OF SELECTION TESTS IN INDUSTRIAL SETTINGS

In industrial manufacturing settings, hiring a production or maintenance worker who lacks the necessary knowledge, skills and abilities to perform the job can be especially costly. In addition to possibly causing bodily injury to themselves or others, unqualified workers threaten to cause production losses, and/or serious damage to equipment, costing hundreds of thousands of dollars. Furthermore, for many companies, errors in work could result in consumer product liability lawsuits costing millions of dollars. When companies are faced with a large numbers of applicants to screen, testing can provide a quick, inexpensive, reliable, and accurate picture of job-related attributes. Two types of tests are commonly used to select production and maintenance workers in industrial settings: multiple-choice (MC) job knowledge tests and performance assessments (PAs). In instances where a MC job knowledge test and PAs are used together in the same selection procedure, the two types of tests are intended to complement one another. Generally, MC job knowledge tests are designed to measure what an applicant *knows* about the job, while the PAs are designed to measure an applicant's ability to *do* or *perform* the job or certain aspects of the job (Callinan & Robertson, 2000; Haladyna, 1994).

MC job knowledge tests differ from PAs in that *knowing* about the job is different than *doing* or *performing* the job (Callinan & Robertson, 2000). *Knowledge* refers to any cognitive behavior of an abstract nature, usually involving such content as facts, concepts, principles, or procedures (Haladyna, 1994). MC exams are designed to measure *knowledge*. Doing or performing a job requires a s*kill*, which refers to the actual performance or the result of a performance (Haladyna, 1994). PAs are designed to demonstrate actual performance of a *skill*.

MC tests are typically recommended for measuring knowledge while PAs are generally recommended for measuring skills. However, complex mental acts such as reasoning, critical thinking, and problem solving may be effectively measured with either format (Haladyna, 1994).

1.2 PROBLEM, SETTING AND ASSESSMENT INSTRUMENT

A large consumer products manufacturing company in the southeastern United States planned to close an older facility and move its production and maintenance workers to a new facility located in the same city. A selection procedure consisting of both MC paper-and pencil and hands-on PA elements was developed to ensure that workers possessed the necessary knowledge, skills, and abilities to perform the jobs at the new facility. Two 125-item alternate tests using an MC format and 7 PA exercises were initially developed for the job, referred to in the present study, as L2/L3 Production Technician.

1.3 ITEM AND TEST ANALYSIS

Item analysis is defined as the computation and examination of any statistical property of an item response distribution (Crocker & Algina, 1986). The purpose of item analysis is to improve test items, which in turn, improves tests and the results of test scores (Wainer, 1989). Indices that are typically provided by a traditional item analysis include measures of item difficulty and discrimination, frequency of response for each option for low and high test scorers, and test score characteristics, including descriptive statistics and reliability (Haladyna, 1994).

1.3.1 Parallel Tests

Employers have to deal with a variety of test security issues and must make certain that examinees do not copy from one another or take the same test form twice in a retesting situation. The use of parallel test forms is one method to address test security issues. Parallel tests consist of two or more test forms that are built according to the same test specifications but feature a different set of test questions (Millman & Green, 1989).

1.4 OVERVIEW OF VALIDITY EVIDENCE

Validity is a unitary concept, which looks to multiple sources of evidence to support the proposed interpretation or use of assessment scores. The process of validation involves collecting evidence to build an argument for the proposed use of the test results. The strength of the evidence determines the degree of validity. The *Standards for Educational and Psychological Testing* (AERA et al, 1999) discuss five distinct sources of validity evidence: content, responses, internal structure, relationship to other variables or external validity, and consequences. Some types of assessment may require a stronger emphasis on one or more sources of evidence compared to other assessment methods.

1.4.1 Content Validity Evidence

Content analysis employing a content specialist or subject-matter expert to evaluate the extent to which the PA exercises measure what is on the MC test, as well as other competencies not covered by the MC test, would provide evidence of the nature of the relationship between the two assessments. Content specialists are persons with in-depth knowledge of the subject-matter who are willing to review items to ensure that each item represents the content and level of cognitive behavior desired (Haladyna, 1994). The use of content specialist judgments to assess the relationship between MC tests and PA exercises offers significant potential as this approach is not dependent on group composition or instructional effects, may not require complicated statistical techniques, is not limited to highly structured content domains, and can be implemented easily in practical settings (Rovinelli & Hambleton, 1976). According to Rovinelli and Hambleton, when utilizing the services of content specialists, one should use the simplest of techniques available to collect data and structure the response task for the content specialist in a way that is neither tedious nor time consuming.

1.4.1.1 Generalizability Theory Generalizability theory (G-theory) is a measurement theory that enables the isolation, and quantification of different sources of variation in a measurement situation using the analysis of variance. With G-theory, for a given measurement situation, sources of error can be isolated and examined, and this information can be used to modify measurement conditions of future studies to maximize reliability. For example, G-theory could be applied to a study with two sources of error: items and raters. If it were determined that raters comprised the greatest source of error, future studies could include ways to reduce rater error such as including more extensive training for the raters.

G-theory can also be used to produce a G-coefficient, which is similar to the reliability coefficient in classical test theory, that is the proportion of expected observed-score variance that is universe-score variance (Shavelson & Webb, 1991). In a study with raters evaluating individual items, the G-coefficient indicates consistency among raters in the same way that the reliability coefficient for raters (intraclass correlation coefficient) does.

1.4.2 Evidence of the Internal Structure of an Assessment

While human judgment by content specialists may be valuable for the confirmation of item content, statistical methods also exist that provide a reliable basis for helping improve tests and the interpretation of test scores. One such technique is factor analysis which assumes that the observed variables are linear combinations of some unobservable, underlying factor (Kim & Mueller, 1978). Factor analysis studies can be characterized as being either exploratory or confirmatory. In exploratory factor analysis (EFA) the objective is not to verify a factor structure but rather to try to find a factor structure that could account for the intercorrelations of an observed set of variables. EFA is a useful technique for investigating the underlying patterns of the data where a compelling theory of the underlying structure of the variables is not readily apparent or in areas where theory is not well established.

Confirmatory factor analysis (CFA) is a means for grouping items into content and/or process categories. It is a useful method that can be used to verify the reasoning that goes into test specifications, providing empirical evidence for the content and/or process categories of a test. In the confirmatory factor model, the researcher must determine in advance of analysis which constraints to impose that determine (a) which pairs of common factors are correlated, (b) which observed variables are affected by which common factors, (c) which observed variables are affected by a unique factor, and (d) which pairs of unique factors are correlated (Long, 1983). Furthermore, CFA allows for statistical tests to be performed to determine if the sample data are consistent with the imposed constraints (Long, 1983).

1.4.3 Evidence of the External Validity of an Assessment

An analysis of the relationship between MC test scores and scores on the PA exercises provides a valuable source of validity information. Evidence of a moderate to strong positive relationship between the two different types of assessments would be consistent with the theoretical framework of the constructs being measured. Evidence based on relationships between the two types of assessments would certainly include correlational evidence. Past research has shown significant positive correlations between various assessment types, such as MC tests and PAs (Breland & Griswold, 1982; Hattrup & Schmitt, 1990; Hogan, Arneson, & Petersons, 1992). Messick (1993) states that, “although in the interest of reality testing and generalizability it would indeed be desirable if the test were related to real-world behavioral variables, what is critical is that it relate appropriately to other construct scores based on distinctly different measurement methods from its own."

1.4.3.1 Linear Regression Linear regression is a statistical technique that attempts to model the relationship between two variables by fitting a linear equation to observed data (Neter, Kutner, Nachtsheim & Wasserman, 1996). One variable is considered to be the independent or predictor variable, and the other is considered to be a dependent or response variable. Regression analysis serves three, frequently overlapping, purposes: description, control and prediction. A regression analysis can be used to provide external validity evidence by describing and predicting the relationship between the MC test and PA exercises.

1.4.3.2 Decision Consistency Estimated with Two Tests Decision consistency refers to the degree to which the same decisions are made from two different sets of measurements (Crocker & Algina, 1986). Decision consistency can be used to evaluate the consistency of decisions based on different types of mastery tests administered to the same group of examinees. Decisions for an examinee are consistent when the results of both tests indicate that an examinee should be classified as passing. A decision consistency analysis provides an additional source of information regarding the relationship between the MC test and PA exercises.

The selection procedure developed for the job of L2/L3 Production Technician was designed to assess necessary knowledge, skills, and abilities with both MC and PA formats. Additional research is needed to examine the relationship between the MC test items and the PA exercises and to provide evidence to support inferences based upon scores for both assessments. This evidence should demonstrate both the internal structure and the external validity of both types of assessment items.

1.5 PURPOSE OF THIS PROJECT AND RESEARCH QUESTIONS

The purpose of this project is to provide evidence of the relationship among items within two alternate MC job knowledge tests and the PA exercises. The sections that follow describe the research questions and the methods, including the procedures for data collection and analysis, to address each research question.

The research questions that will be investigated in this study:

1. Based on a content analysis of the assessments, what knowledge, skills, and abilities are measured by the MC job knowledge test and the PA exercises?
    - 1a. Based on a content analysis of the assessments, to what extent do the PA exercises measure the same knowledge, skills, and abilities as the MC job knowledge test items?
    - 1b. Based on a content analysis of the assessments, what additional knowledge, skills, or abilities are being assessed by the PA exercises beyond what is measured by the MC job knowledge test items?
    - 1c. Based on the results of the content analysis of the assessments, which subtests of the MC job knowledge tests are most related to each of the seven PA exercises?
2. What are the item and test properties of the MC job knowledge test items and the PA exercises?
    - 2a. To what degree are item and test properties similar for both Form A and Form B?
3. What is the factor structure of both Form A and Form B with the PA exercises included?

4. What is the relationship between scores on the MC job knowledge tests and total score on the PA exercises?

    4a. Based on a regression analysis, what is the nature of the relationship between the predictor variable MC test score and the dependent variable PA test score?

    4b. What is the decision consistency regarding those who score above and below the cut score on the MC test and the PA exercises?

5. What is the relationship between each subtest of the MC job knowledge test and each of the seven PA exercises?

2.0 REVIEW OF THE LITERATURE

2.1 TESTING FOR SELECTION

In some cases, the hiring of the wrong person for a job can cost a company millions of dollars depending upon the size of the organization and the importance of the job in question. In instances where there are large numbers of applicants to screen, tests provide a quick inexpensive, reliable, and in most cases a more accurate picture of job-related attributes than other assessment methods such as interviews or graphoanalysis, which is a system of handwriting analysis used to detect personality traits (Ramsay, 2003). Furthermore, testing is often easier to defend legally than other less documentable means (Ramsay, 2003). The present study involves two types of tests commonly used to select individuals in industrial settings, multiple-choice job knowledge tests and performance assessments.

2.1.1 Job Knowledge Tests

Job knowledge tests are most often paper-and- paper measures of the amount of information an examinee possesses about a job (Callinan & Robertson, 2000). Job knowledge tests have been shown to be good predictors of job knowledge. Robertson and Kandola (1982) reported a median validity of .4 for predicting job performance for job-related information tests. Schmidt and Hunter (1998) calculated the predictive validity of job knowledge tests at .48. Carey (1991) found that a job knowledge test for US Marines was a suitable substitute for hands-on performance tests in personnel selection.

The multiple-choice format has many desirable features over other formats in the areas of: ease of item and test construction, administration, scoring, analysis and evaluation of test items, guessing, reliability, and validity (Haladyna, 1994).

MC job knowledge tests differ from PA measures in that *knowing* about the job is different than *doing* or *performing* the job (Callinan & Robertson, 2000). According to Haladyna (1994) *knowledge* refers to any cognitive behavior of an abstract nature, typically involving such content as facts, concepts, principles, or procedures. Written tests such as multiple-choice exams are designed to measure *knowledge and the application of knowledge.* Doing or performing a job requires a *skill*, which refers to the actual performance or the result of a performance, namely a product (Haladyna, 1994). Performance assessments are considered to be constructed-response exercises designed to demonstrate actual performance of a *skill* (Haladyna, 1994).

Although MC tests are typically recommended for measuring knowledge and constructed-response formats are generally recommended for measuring skill, complex mental acts such as reasoning, critical thinking, and problem solving may be effectively measured with either format (Haladyna, 1994).

2.1.2 Performance Assessment

Performance Assessment (PA) is defined as a procedure which requires examinees to complete tasks or processes that demonstrate their ability to apply knowledge and skills, or to put knowledge and understanding into action in simulated or real-life situations (Messick, 1996; Nitko, 1996).

PA is considered to be highly suitable for the assessment of higher-order thinking or problem-solving skills. PA allows the structure of responses to be defined by the examinee, resulting in the capability to score for multiple levels of quality versus only as correct or incorrect. Accordingly, PA can demonstrate skills that are not easily assessed with multiple-choice (MC) items (Messick, 1996). PA is also seen by many (e.g. examinees, educators, and teachers) as more authentic than traditional MC items. The perceived authenticity of PA over MC items is likely due to the fact that PA emphasizes problem solving, reasoning, and the ability to integrate knowledge and information, rather than only providing isolated bits of knowledge and information (Muraki, Hombo, & Lee, 2000).

In the realm of employment testing performance tests are often called "work sample tests". As in the previous definition of PA, a work sample test is a hands-on performance test in which a job applicant or employee is required to actually perform a job-related task under the same conditions as those required on the job (Callinan & Robertson, 2000). Work samples are typically used as predictor measures for the purpose of personnel selection.

When compared with other selection methods, some types of work sample tests have demonstrated higher predictive validity than general mental ability. Schmidt and Hunter (1998) in a meta-analytic study of 19 selection procedures in predicting job performance found the highest reported validity for an individual method was for work sample tests.

Work sample tests also appear to have substantially less adverse impact against minority groups. Schmidt, Clause and Pulakos (1996), found little to no difference between African-

American and White applicants and Hispanic-American and White candidates on work sample performance.

Job-relatedness, or content validity evidence, is an important issue concerning the development of work sample tests. Work sample tests usually receive positive reaction from applicants as they are seen as job-related and therefore are perceived to be fair (Steiner & Gilliland, 1996). Work sample tests also function as a realistic job preview as well as a selection tool because they reflect aspects of the actual job (Downs, Farr, & Colbeck, 1978).

Work sample testing is not without its limitations (Callinan & Robertson, 2000). Work sample tests are not appropriate for assessing applicants without job experience because they require specific procedural job knowledge. Furthermore, development and administration can also be time consuming and costly as work sample tests are typically administered individually in the actual workplace or in a specially constructed simulated context. Work sample or PA exercises are also: (1) often multidimensional and unstable across contexts, (2) typically feature fewer items than MC tests, resulting in very unstable scores, (3) easy to memorize and as such are not reusable, and (4) often complex to score due to their typically polytomously scored nature (Muraki, Hombo, & Lee, 2000).

2.2 ITEM AND TEST ANALYSIS

Item analysis is defined as the computation and examination of any statistical property of an item response distribution (Crocker & Algina, 1986). The three general types of indices that are typically obtained from an item analysis (a) serve to describe the distribution of responses to a single item, (b) serve to describe the degree of relationship between response to the item and a criterion, or (c) are a function of both item variance and relationship to a criterion.

2.2.1 Item Difficulty

For tests consisting of dichotomously scored items, the mean item score corresponds to the proportion of examinees (p_i) who answered the item (i) correctly. Item difficulty or (p_i) can range from .00 to 1.00. While not an indicator of item quality, multiple-choice items with p values slightly above .50 will allow total score variance, and consequently reliability, to be maximized (Crocker & Algina, 1986).

2.2.2 Item Discrimination

Item discrimination indices serve as a measure of how effectively the item discriminates between examinees that are relatively high on the criterion of interest and those who are relatively low (Crocker & Algina, 1986). When an internal criterion is used (e.g. total test score) the goal is to identify items for which high-scoring examinees have a high probability of answering correctly and low-scoring examinees have a high probability of missing. Items that are missed equally by high and low scoring examinees are undesirable. In particular, items that are missed by high scoring examinees but answered correctly by low scoring examinees indicate potentially bad items.

2.2.3 Point Biserial Correlation

While there are several different indicators of an item's discrimination effectiveness, one of the most commonly used involves a correlation coefficient, called the point biserial correlation. For items that are scored 0 to 1 the point biserial correlation can be used to determine how closely performance on the test item is related to performance on the total test or test section. The formula for the point biserial correlation is,

$$\mathrm{P}_{pbis} = \frac{(\mu_{+} - \mu_{x})}{\sigma_{x}} \sqrt{p/q} \qquad (1)$$

where, μ_{+} is the mean criterion score for those who answered the item correctly, μ_{x} is the mean criterion score for the entire group, σ_{x} is their standard deviation, *p* is the item difficulty, and *q* is (1-*p*).

For dichotomously scored items, Lord and Novick (1968) recommend the point biserial correlation if it is expected that future samples of examinees will be similar in ability to the item analysis sample, and the goal is to select items that will have high internal consistency.

2.2.4 Evidence for Internal Consistency of the Test

A reliability study that involves the administration of a single form of a test to a group of examinees is concerned with the internal consistency of the test. Analysis of data in such a study yields a coefficient which provides an estimate of how consistently examinees perform across items within a test during a single testing session (Crocker & Algina, 1986).

One method of evaluating the internal consistency of a test is to separately score two halves of a test for each examinee. The half-test scores are then correlated and corrected with the Spearman Brown formula, or the difference between half-test scores could be computed and the reliability estimated using Rulon’s method (Crocker & Algina, 1986). Rulon’s method uses the difference score between the half-tests $D = A\text{-}B$ where A is the examinee's score on the first half-test and B is the score on the second half test. The formula for Rulon's method is:

$$\rho_{XX'} = 1 - \frac{\hat{\sigma}_D^2}{\hat{\sigma}_x^2} \qquad (2)$$

where σ_D^2 is the variance of the difference scores, and σ_x^2 is the variance of the observed scores (Crocker & Algina, 1986).

The major inadequacy of split-half procedures is that different ways of splitting the test result in different reliability estimates. One formula that addresses the problem of split-half techniques is coefficient alpha. Coefficient alpha is the average of all the split-half coefficients that would be obtained if the test were divided into all possible half-test combinations and the reliability estimated by using Rulon's procedure (Crocker & Algina, 1986).

Coefficient Alpha is computed by the formula,

$$\hat{\alpha} = \frac{k}{k-1}\left(1 - \frac{\Sigma\hat{\sigma}_i^2}{\hat{\sigma}_x^2}\right) \qquad (3)$$

where k is the number of items on the test, $\hat{\sigma}_i^2$ is the variance of item *i*, and σ_x^2 is the total test variance (Crocker & Algina, 1986).

Several factors in the testing situation can have an impact on obtained reliability estimates (Crocker & Algina, 1986). Group homogeneity affects reliability estimates in that coefficients will be lower for groups highly homogeneous on the measured trait than for groups that are more heterogeneous. Speeded tests may produce artificially inflated test reliability coefficients because uncompleted test items will be perfectly consistent in spite of differences in item content. Finally, test length affects reliability estimates such that longer test are more reliable than shorter tests composed of similar items. The effect of varying test length can be estimated by means of the Spearman Brown prophecy formula.

2.2.5 Parallel Test Forms

The use of parallel test forms is one method to address test security issues. Parallel tests consist of two or more test forms that are built according to the same test specifications but feature a different set of test questions (Millman & Green, 1989).

If the different test forms differ somewhat in difficulty, then a statistical test process known as equating can be used to adjust scores on test forms so that scores on the forms can be used interchangeably (Kolen & Brennan, 1995).

Harris and Crouse (1993) identify four conditions for equating that they attribute to Lord (1980):

1. The test to be equated must measure the same construct.
2. The conditional distributions of scores given the true score on each test after equating must be equal (this is termed equity).
3. The equating transformation should be invariant across populations.
4. The equating transformation should be symmetric. (p. 196)

According to Lord, (1980) equity as it applies to the current study, means that it does not matter to each examinee whether they take Form A or Form B. When the two tests are perfectly parallel, the equity property will hold making equating unnecessary.

2.3 OVERVIEW OF VALIDITY EVIDENCE

According to the current *Standards for Educational and Psychological Testing*: "Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (American Educational Research Association [AERA], American Psychological Association [APA], and National Council on Measurement in Education [NCME], 1999, p.9). The previous conceptualization of validity was defined as three separate types: content, criterion and construct. The contemporary model views validity as a unitary concept, which looks to multiple sources of evidence to support the proposed interpretation or use of assessment scores. The process of validations involves collecting evidence to build an argument for the proposed use of the test results. The strength of the evidence determines the degree of validity.

The *Standards for Educational and Psychological Testing* (AERA et al, 1999) discuss five distinct sources of validity evidence: content, responses, internal structure, relationship to other variables sometimes referred to as external validity evidence, and consequences. Some types of assessment may require a stronger emphasis on one or more sources of evidence compared to other assessment methods.

2.4 SOURCES OF VALIDITY EVIDENCE

2.4.1 Evidence Based on Test Content

Test content according to the *Standards for Educational and Psychological Testing,* "refers to the themes, wording, and format of the items, tasks, or questions on a test, as well as the guidelines for procedures regarding administration and scoring" (AERA et al, 1999, p.11).

Evidence based on test content refers to how relevant and representative the test content is to the domain or universe of interest. Validity evidence based on test content might be obtained from expert judgments of the extent of the relationship between a test's content and the construct it is intended to measure for example.

2.4.2 Evidence Based on Response Processes

Evidence based on response processes involves information that the processes used by examinees in responding to a test are those that the test was actually intended to assess (AERA et al, 1999). For example, evidence based on response processes may involve asking examinees to think aloud while they are attempting to answer a question to determine if the intended construct is indeed being assessed. Validity evidence based on response process may also include information related to, examinee format familiarity, quality control of scoring, or accuracy of applying pass-fail decision rules to scores (Downing, 2003).

2.4.3 Evidence Based on Internal Structure

Evidence based on internal structure refers to the extent to which the statistical or psychometric characteristics of the test questions and/ or parts of the test correspond to the construct hypothesized to underlie test performance (AERA et al, 1999). For example, a test of a construct conceptualized as having a unidimensional structure should show evidence of item homogeneity. Evidence based on the internal structure of the test may also include item difficulty/discrimination, item/test characteristic curves, inter-item correlations, item-total correlations, score scale reliability, or standard errors of measurement (Downing, 2003).

2.4.4 Evidence Based on Relations to Other Variables

Analysis of the relationship of test scores to variables external to the test provide evidence on whether scores on the test may indeed be interpreted in terms of the construct intended. (AERA et al, 1999) This type of evidence may seek higher positive correlational evidence with criteria hypothesized to measure the same construct (convergent evidence) or lower correlational evidence with measures purportedly of different constructs (discriminant evidence). The other variables may include criteria that will become available some time in the future (predictive designs) or are currently available (concurrent designs). Criterion variables should reflect attributes that are of primary interest to the researcher.

In some circumstances, there may be a strong basis to generalize test-criterion relationships from one situation to another. However, limitations in a study's design, missing data, or a lack of variance in scores for example, may limit the generalizability of the validity coefficients beyond the original study.

2.4.5 Evidence Based on Consequences of Testing

The consequential aspect of validity refers to the intended or unintended, positive or negative, impact that results from the use of an assessment (AERA et al, 1999). The consequences of testing include decisions, outcomes, and interpretations based on the assessment scores. Evidence based on the consequences of testing may include the impact of test scores on society, the consequences on future learning, and instructional/learner consequences (Downing, 2003).

2.5 CONTENT VALIDITY EVIDENCE

2.5.1 Content Analysis

Content analysis when used as part of test validation has typically employed content specialists or subject-matter experts (SMEs) to evaluate the content of the assessment. Content specialist are persons intimate with the subject-matter who are willing to review items to ensure that each item represents the content and level of cognitive behavior desired (Haladyna, 1994). The use of content specialists to evaluate test information can serve different functions. Hambleton and Rogers (1988), for example, designed a review form to be completed by content specialists to aid in detecting item bias in tests. Many studies have also used content specialists to evaluate the instructional or content domain representation of a test or assessment (Hemphill & Westie, 1950; Rovinelli & Hambleton, 1976; Dolmans, Gijselaers & Schmidt, 1992; Sireci & Geisinger, 1995). Since all inferences derived from test scores are valid only to the extent to which the test measures the constructs it claims to measure, content domain representation is vital for demonstrating the validity of inferences derived from test scores (Yalow & Popham, 1983; Sireci, 1995; Sireci & Geisinger, 1995).

The methods utilized during the content-related validation study for the development of the L2/L3 Production Technician MC and PA exercises represent an *a priori* approach to providing validity evidence by developing a direct relationship between an item and test objective or blueprint during the item/test construction phase. The procedures described in the remainder of this section represent *a posteriori* procedures which are designed to assess whether or not a direct relationship between an item and an objective exists through the analyses of data conducted after the item is written. While the *a posteriori* methods described here were developed for use in assessing instructional objective representation, they can also be seen as useful techniques for assessing the relationship between the two types of L2/L3 Production Technician assessments.

Rovinelli and Hambleton (1976) reviewed three methods for the collection and analysis of content specialist judgment data: the Semantic Differential Technique, the Matching Procedure, and the Hemphill-Westie procedure. While their study was done in the context of assessing item validity, their comparison and analysis of possible data collection techniques and methods of analyzing content specialists' ratings are relevant to the evaluation of the L2/L3 Production Technician assessments (Rovinelli & Hambleton, 1976).

Rovinelli and Hambleton (1976) identify five questions regarding the use of content specialists' ratings:

1. Can the content specialists make meaningful (valid) judgments about the relevance of items to instructional content?
2. Is there agreement amongst the ratings of content specialists?
3. What information is one seeking to obtain from the judgmental data?
4. What variables affect the judgmental techniques?
5. What techniques can be used for collecting content specialists' ratings of test items? (p.7)

2.5.1.1 Semantic Differential Technique A frequently used procedure for the collection and analysis of content specialist judgment data involves the use of the semantic differential procedure (Rovenelli & Hambleton, 1976). The content specialists are presented with a PA exercise as well as all of the MC items for which ratings are desired. They are asked to make a judgment which consists of deciding whether the relationship between the MC item and the PA exercise is best described by the adjective toward the left end or right end of the scale. A semantic differential scale might look like this: (a) No Relationship, (b) Small Relationship, (c) Moderate Relationship, (d) Strong Relationship.

One advantage of this technique is that obtained data can be analyzed without employing sophisticated statistical procedures. However, the data also lends itself to more elaborate statistical analysis if necessary. If several content specialists are involved, an examination of the standard deviations of the scores can be used to provide an indication of the extent of agreement.

Aiken's (1980) validity index accounts for the number of categories used to rate each item and the number of judges that respond to each category. Aiken's validity index, V is given by:

$$V = \frac{\sum_{i=1}^{c-1} i n_i}{N(c-1)'} \qquad (5)$$

where c is the number of categories on the item relevance rating scale, i is the weight given to each category, n_i is the number of judges who rated the item in the ith category, and N is the total number of subject-matter experts.

The lowest category is given a weight (or i-value) of 0, the next category is given a weight of 1, and so on, and the highest category is given a weight of c-1 (Sireci & Geisinger, 1995).

Hambleton (1984) suggested that relevance or relationship data be averaged over the number of content specialists and the mean relevance rating for each item on each criterion, such as a PA exercise, be computed.

2.5.1.2 Matching Procedure Another procedure used to obtain the judgments of content specialists involves the use of a matching task (Rovenelli & Hambleton, 1976). The content specialists are presented with two lists. The first list contains the set of MC items and the second list is a set of the PA exercises. The content specialist matches or assigns items to the PA exercises that they feel measure some aspects of the knowledge, skill or ability. Rovinelli and Hambleton (1976), suggest that a contingency table can be constructed to represent the number of times each item is assigned to each PA exercise across the content specialists. While statistical tests can be performed on the results, a simple visual inspection of the contingency table will provide information on which MC items measure some aspects of the same knowledge, skills and abilities as the PA exercises. The matching procedure was used by Dolmans, Gijselaers, and Schmidt (1992) who used content specialists to assess the overlap between the intended curricular content and the information required to answer achievement test items correctly. Teachers who served as content specialists were asked a posteriori to assign test items to one or more of the topics presented in a topic list of curriculum content (Dolmans et al., 1992).

Sireci and Geisinger (1995) employed the matching procedure with groups of subject-matter experts to evaluate the content domain representation of a national licensure examination and a nationally standardized social studies achievement test. The SMEs rated the relevance of the items to the content domains listed in the test blueprints. Two methods of assessing content representation were used: (a) Multidimensional scaling (MDS) and (b) the item relevance ratings were analyzed using procedures proposed by Hambleton (1984) and Aiken (1980). The results of the MDS solutions agreed with the subject-matter experts' perceptions of the underlying content structure of the tests.

2.5.1.3 Index of Item Homogeneity In 1950, Hemphill and Westie devised an index of homogeneity of placement for use in constructing personality tests. The Index of Item Homogeneity is a numeric representation of the judgment of content specialists on the degree to which they feel that an item belongs to one unique personality dimension. By substituting "PA exercise" for "personality dimension", the Index of Item Homogeneity can be used to evaluate the relationship between MC items and PA exercises.

The Index of Item Homogeneity consists of having the content specialists rate each MC test item on each of the PA exercises by assigning a value of +1, 0, or -1 where, (a) +1 = definite feeling that an MC item is a measure of some aspects of a PA exercise, (b) 0 = undecided about whether the MC item is a measure of some aspects of a PA exercise, and (c) -1 = definite feeling that an MC item is not a measure of some aspects of a PA exercise. Hemphill and Westie's (1950) formula for the Index of Homogeneity is as follows:

$$I_{ij} = \frac{N \sum_{j=1}^{n} X_{ijk} - \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ijk}}{2 \cdot 2n(N-1) + \sum_{i=j}^{N} \sum_{j=1}^{n} X_{ijk} - \sum_{j=1}^{n} X_{ijk}} \tag{6}$$

where I_{ik} is the Index of Homogeneity for item k on PA exercise i, N is the number of PA exercises (i=1, …, N), n is the number of content specialists (j=1,…, n) and X_{ijk} is the rating (1, -1, or 0) of item k as a measure of PA exercise i by content specialist j (Hemphill & Westie, 1950).

The Hemphill-Westie procedure is not without its shortcomings, however (Rovenelli & Hambleton, 1976). First, the minimum and maximum values are .67 and -.40 respectively. The maximum value of this index will occur when each content specialist assigns a +1 to the item for the appropriate PA exercise and a -1 for all the other PA exercises. The minimum value will occur when content specialists assign a -1 to the item for the appropriate PA exercise and a +1 for all the other PA exercises. As a result, interpreting the Index of Homogeneity is more difficult than if the index ranged from -1 to +1. Second, and even more problematic, is that the index's value varies as a function of the number of content specialists and PA exercises, which complicates the interpretability of the index (Rovenelli & Hambleton, 1976).

Given the deficiencies of the Hemphill-Westie formula, Rovinelli and Hambleton (1976) developed a new formula called the Index of Item-Objective Congruence. The Index of Item-Objective Congruence has the following assumptions (Rovenelli & Hambleton, 1976):

1. That perfect item objective congruence should be represented by a value of +1 and will occur when all of the specialists assign a +1 to the item for the appropriate objective and a -1 to the item for all the other objectives.
2. That the worst judgment an item can receive should be represented by a value of -1 and will occur when all the specialists assign a -1 to the item for the appropriate objective and a +1 to the item for all the other objectives.
3. That the assignment of a 0 to an item is poorer than a +1 but better than a - 1. This is in effect saying that it is better for a specialist to not be able to definitely decide whether an item is a measure of an appropriate objective than it is for the judge to feel that the item is definitely not a measure of the objective.
4. That this index should be invariant to the number of content specialists and the number of objectives (p. 15).

Substituting the term "PA exercise" for the term "objective" the formula for the Index of Item-Objective Congruence is as follows:

$$I_{ik} = \frac{(N-1)\sum_{j=1}^{n} X_{ijk} - \sum_{i=1}^{N}\sum_{j=1}^{n} X_{ijk} + \sum_{j=1}^{n} X_{ijk}}{2(N-1)n} \tag{7}$$

where I_{ik} is the Index of Homogeneity for item *k* on PA exercise *i*, *N* is the number of PA exercises (i=1, …, *N*), *n* is the number of content specialists (*j*=1,…, *n*) and X_{ijk} is the rating (1, -1, or 0) of item *k* as a measure of PA exercise *i* by content specialist j (Rovenelli & Hambleton, 1976).

Similar to the Hemphill-Westie Index, the Index of Item-Objective Congruence offers no method for determining the statistical significance of the values for the Index of Item-Objective Congruence. However, the use of Lu's coefficient of agreement (1971) amongst the content specialists can be used to give an indication of how consistent the judgments are (Rovenelli & Hambleton, 1976).

Of the three procedures reviewed by Rovinelli and Hambleton (1976), the Hemphill-Westie procedure was recommended over the other two techniques. Two reasons are offered for support of the Hemphill-Westie procedure: (a) the numeric representation of the data aids in interpretation, (b) there are means for determining the reliability and validity of the data collected, and these methods can be tested for significance (Rovinelli & Hambleton, 1976). However, the Hemphill-Westie procedure is not without its limitations (Rovinelli & Hambleton, 1976). First, the procedure cannot be used to collect information of such topics as item and distracter quality. Second, the dimensionality of the data must be known in advance of its use. Third, the Hemphill-Westie procedure is quite time consuming with large numbers of items.

2.5.1.4 Conclusions on Content Analysis The use of content specialists' judgments to assess the relationship between MC and PA exercises offers significant potential as this approach is not dependent on group composition or instructional effects, may not require complicated statistical techniques, is not limited to highly structured content domains, and can be implemented easily in practical settings (Rovinelli & Hambleton, 1976). According to Rovinelli and Hambleton (1976), when utilizing judgmental procedures, one should use the simplest of techniques available to collect data and structure the response task for the content specialist in a way that is not tedious or time consuming.

2.5.1.5 Generalizability Theory A person's universe score (true score) is considered his/her score on all admissible observations. The extent to which the sample of admissible observations allows the estimate of the true score determines the generalizability of the measurement (Shavelson & Webb, 1991). Generalizability theory (G-theory) is a measurement theory that enables the isolation, and quantification of different sources of variation in a measurement situation using the analysis of variance. With G-theory, for a given measurement situation, sources of error can be isolated and examined, and this information can be used to modify measurement conditions of future studies to maximize reliability. For the present study, with two sources of error: items and raters, if G-theory determined that raters comprised the greatest source of error, future studies could include ways to reduce rater error such as including more extensive training for the raters, or increasing the number of raters.

As with classical test theory (CTT), G-theory assumes a person's observed score is comprised of his/her universe score (true score), and one or more sources of error (Shavelson & Webb, 1991). In G-theory, errors are assumed to be independent of true scores and uncorrelated, and the samples used to estimate the error variances consist of random samples from their particular populations.

Whereas, classical test theory (CTT), has a reliability coefficient to inform about a single source of measurement error, G-theory informs about error due to multiple sources of error at once. In G-theory, these multiple sources of measurement error are called *facets* and reflect different sources of variations (Shavelson & Webb, 1991). Facets might include items, occasions, raters, or locations for example. In the social sciences, *persons* is usually *the object of the measurement*. However, in the present study, which features a one-facet design for raters evaluating MC items and their relationship to a particular PA exercise, *items* are the object of measurement, and there are four sources of variability, (a) differences among the objects of measurement (items), (b) differences among raters, (c) differences in the item-by-rater match, and (d) random or unidentified events (Shavelson & Webb, 1991).

In G-Theory, facets can be treated as *random* or *fixed* (Shavelson & Webb, 1991). Facets are considered random if the sample is considered to be interchangeable with any other sample of the same size drawn, from a much larger universe. Facets are treated as fixed when they reflect the conditions of the entire population (or only ones of interest).

G-theory studies can have either *crossed* or *nested* designs (Shavelson & Webb, 1991). With crossed designs, all units of one facet are associated with all units of another facet, for example, persons x raters x items (p x r x i).   Nested designs feature each set of units from one facet associated with a unique unit from another facet, for example, persons could be nested within raters and would be indicated by (p:r).

G-theory also distinguishes between *relative* and *absolute* decisions (Shavelson & Webb, 1991). Relative decisions are those used to compare individuals to each other, while absolute decisions are those based on an individual's absolute level of performance. For relative decisions, all variance components include the interactions of each facet with the object of measurement, while measurement error for absolute decisions include all variance components except the object of measurement.

The G-coefficient, which is similar to the reliability coefficient in classical test theory, is the proportion of expected observed-score variance that is universe-score variance (Shavelson & Webb, 1991). In a $p \times r$ design, persons (p) are the targets of measurement, and rater (r) is treated as a random facet. In the present study, MC items (mc_item) are the targets of measurement and the intent is to generalize the measurement across the random facet: raters (rater). The formula for calculating a G-coefficient for a $p \times r$ design is:

$$\rho^2 = \frac{\sigma^2{}_p}{\sigma^2{}_p + \left[\frac{\sigma^2{}_r}{n_r} + \frac{\sigma^2{}_{pr}}{n_r n_i} \right]} \tag{8}$$

In the present study with raters evaluating individual items, the G-coefficient indicates consistency among raters in the same way that the reliability coefficient for raters (intraclass correlation coefficient) does.

2.6 INTERNAL VALIDITY EVIDENCE

2.6.1 Overview of Exploratory Factor Analysis

Variables of theoretical interest that cannot be directly observed are referred to as either latent variables or factors. Even though latent variables cannot be directly observed, information about them can be obtained indirectly by noting their effects on observed variables. Factor analysis is a statistical procedure for revealing a (typically) smaller number of latent variables by studying the covariation among a set of observed variables.

In terms of test data, the item responses would be used to determine which of the items tend to correlate with each other, thus forming a factor structure. The researcher then attempts to describe the resulting factors. A factor structure is not verified in this analysis, but rather sought out. In an exploratory factor (EF) analysis, the researcher does not specify the structure of the relationships among the variables in the model beyond the specifications of the numbers of common factors and observed variables to be analyzed.

When conducting an EFA the researcher must choose a factor extraction method such as weighted least squares, unweighted least squares, generalized least squares, maximum likelihood, or principal axis factoring. Factor indicators for EFA may be continuous, categorical (binary or ordered polytomous), or a combination of continuous and categorical. When at least one factor indicator is categorical, as it is for the L2/L3 Production Technician test data, extraction methods such as weighted least squares (WLS) or unweighted least squares (ULS) are appropriate. In EFA, although one or more of the observed variables may be categorical, any latent variables in the model are assumed to be continuous (Muthen & Muthen, 1998). Additionally, sample size requirements are somewhat more stringent for categorical variables than for continuous variables with at least 200 cases typically required (Muthen & Muthen, 1998).

After extraction the researcher must decide how many factors to retain for rotation. Extracting too many or too few factors retained for rotation can have a detrimental effect on the results. One commonly used approach is to retain only factors with eigenvalues greater than 1 which is known as the Kaiser criterion (Kaiser, 1960). However, the general consensus in the literature is that the Kaiser criterion is one of the least accurate methods for selecting the number of factors to retain as it often retains too many factors (Velicer & Jackson, 1990). Regarding factor loadings, Gorsuch (1983) reports that an absolute value of .3 is commonly used as the minimum loading for interpretation.

An analysis of the graph of the eigenvalues or scree plot is another method for selecting the number of factors to retain (Cattell, 1966). The scree test involves visually inspecting the graph of the eigenvalues and looking for the natural bend or break point in the data where the curve flattens out. The number of datapoints above the break, not including the point at which the break occurs is usually the number of factors to retain.

The next step of EFA is to rotate the initial factor loadings in an attempt to find the simplest and most easily interpreted factor structure. Typical orthogonal rotation methods include varimax, quartimax, and equamax. Common oblique methods of rotation include direct oblimin, quartimin, and promax. Orthogonal rotations produce factors that are uncorrelated while oblique methods allow the factors to correlate. In the social sciences it is expected that there will be some correlation among factors and therefore oblique rotation methods are preferred. If the factors are truly uncorrelated, orthogonal rotation and oblique rotation produce nearly identical results. The final step of EFA is to attempt to interpret or explain the factor structure.

2.6.2 Overview of Confirmatory Factor Analysis

According to Long (1983), in exploratory factor analysis the researcher must assume that (a) all common factors are correlated (or in some cases that all common factors are uncorrelated), (b) all observed variables are directly affected by all common factors, (c) unique factors are uncorrelated with one another (d) all observed variables are affected by a unique factor, and (e) all common factors are uncorrelated with all unique factors. These assumptions are made regardless of the substantive appropriateness. Additional and generally arbitrary assumptions must then be imposed in order to estimate the model's parameters (Long, 1983).

The restrictions of the exploratory factor model have been, for the most part, overcome by the development of the confirmatory factor model (CFA) (Joreskog, 1967, 1969). In the confirmatory factor model, the researcher imposes substantively motivated constraints. According to Long (1983), these constraints determine (a) which pairs of common factors are correlated, (b) which observed variables are affected by which common factors, (c) which observed variable are affected by a unique factor and (d) which pairs of unique factors are correlated. Furthermore, CFA allows that statistical tests be performed to determine if the sample data are consistent with the imposed constraints (Long, 1983).

Since the number of latent variables or factors and the relationships among the factors must be *specified* in advance, CFA should be used when the researcher has some knowledge of the relationships among the data, either through theoretical knowledge or past experience.

A one factor confirmatory factor analysis model is depicted in Figure 1 on the following page. The circle at the top of Figure 1 corresponds to the latent variable ξ_1 with $x_1, x_2, \ldots, x_p$ representing the observed or indicator variables. The $\lambda_1, \lambda_2, \ldots, \lambda_p$, are the factor loadings of the p observed or indicator variables on the latent variable. $\delta_1, \delta_2, \ldots, \delta_p$ are called unique factors or errors in variables. For the L2/L3 Production Technician tests the individual items are the observed or indicator variables.

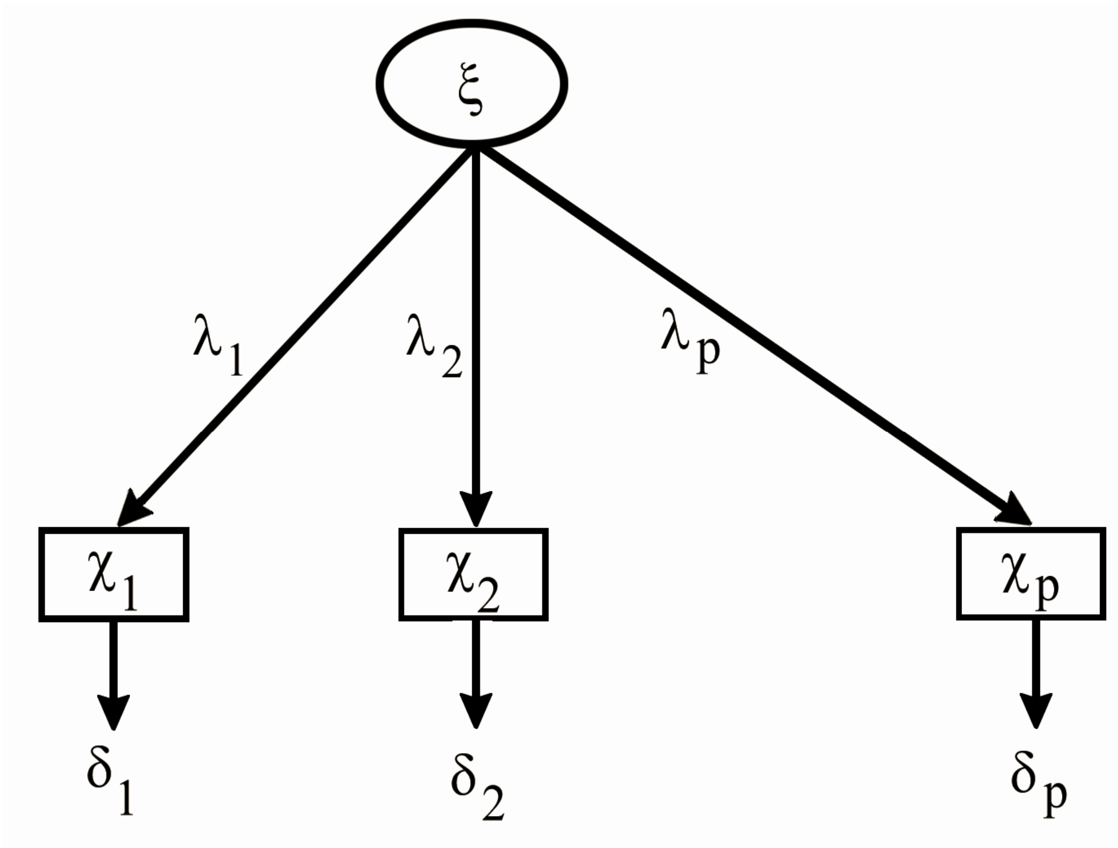

Figure 1. One Factor Model

In confirmatory factor analysis the statistical model reflecting the underlying structure of the data must be specified. The mathematical relationship between the observed variables and the factors is $\mathbf{x} = \Lambda\xi + \delta$. where x is a (q × 1) vector of observed variables, ξ is a (s × 1) vector of common factors, Λ is a (q × s) matrix of factor loadings relating the observed x's to the latent ξ, and δ is a (q × 1) is a vector of the residual or unique factors (Long, 1983). The expected covariance matrix of the observed variables x is given by $\Sigma = \Lambda\Phi\Lambda' + \Theta$, where Φ is a (s × s) covariance matrix of the common factors, Θ is a (q × q) covariance matrix of the residual factors, Λ is again a (q × s) matrix of factor loadings relating the observed x's to the latent ξ, and Λ' is the transpose of the matrix Λ (Long**,** 1983).

The assumptions to be satisfied are (a) variables are measured from their means: $E(\xi) = 0$; $E(x) = E(\delta) = 0$, (b) the number of observed variables is greater than the number of common factors, and (c) common factors and unique factors are uncorrelated: $E(\xi\delta'') = 0$ or $E(\delta\xi'') = 0$ (Long, 1983).

In factor analysis, estimation involves using sample data to make estimates of population parameters (Long, 1983). CFA uses a sample matrix of covariances termed **S,** to estimate the parameters in Λ, Φ and Θ. The estimate of the population covariance matrix Σ, is defined by $\hat{\Sigma} = \hat{\Lambda}\hat{\Phi}\hat{\Lambda}' + \hat{\Theta}$, where the ^ signifies that the matrices contain estimates of population parameters. The null hypothesis that is tested in CFA is that the population variance-covariance matrix of the observed variables is equal to the covariance matrix implied by the specified model. Estimation entails finding values of $\hat{\Lambda}$, $\hat{\Phi}$ and $\hat{\Theta}$, that produce an estimated covariance matrix $\hat{\Sigma}$ that is as close as possible to the sample covariance matrix **S** (Long, 1983).

The weighted least square (WLS) estimation technique is an iterative process that can be implemented in Mplus. WLS uses a weight matrix reflecting the variance/covariance matrix for the sample variance/covariance or correlation matrix. WLS estimator requires a very large sample size. Weighted least squares means and variances adjusted (WLSMV) is a refinement of the WLS estimator and an attempt to reduce the large sample size requirements of WLS by using means and variances only. In order for the WLSMV estimator to produce a unique solution the specified CFA model must be identified. Identification has to do with whether the parameters of the model are uniquely determined. Identification must be established before attempts are made to estimate a model (Long, 1983). If a unique solution to the CFA equation exists, then the model is identified (Comrey & Lee, 1992) One condition that is necessary but not sufficient for model identification is that the number of observed variables, that is, variances and covariances of the observed data, must be greater than the number of parameters that are to be estimated. Constraints can be placed on some parameters to ensure model identification. Values of certain factor loadings or covariances can be set to zero so that they will not be estimated in the model. Thorough consideration of the relevant underlying theory of the problem should be used to determine when factor loadings or covariances are to be fixed.

When too few indicators exist for one or more of the latent variables in the model identification may be difficult to achieve (Loehlin, 1998). According to Bollen (1989), for a one factor model having at least three indicators with nonzero loadings is a sufficient condition to ensure identification. Model identification can be ensured if there are two indicators for each latent variable for models with more than one factor. Identification problems resulting from too few indicators are uncommon with four or more indicators per factor.

After specifying a model and estimating the parameters, the next step is assessing the fit of the model to the data. No one overall best measure of fit has been discovered. Consequently, more than one test or index of fit should be used to assess the fit of a model structure. One of the most popular measures of fit is the χ^2 test, which tests the null hypothesis that the difference between the estimated covariance and sample covariance matrices is zero. A nonsignificant test statistic provides evidence that there is not a gross lack of fit of the model. The usefulness of the χ^2 test is limited by its assumptions that (a) the observed variables are normally distributed, (b) the analysis is based on a sample covariance matrix not a sample correlation matrix, and (c) the sample size is large enough to justify the asymptotic properties of the chi-square test (Long, 1983). At least one of these assumptions is generally violated when confirmatory factor analysis is conducted. Because the χ^2 test is sensitive to sample size, even small differences between the hypothesized and observed structures will lead to rejection of the null hypothesis. Consequently, it is common to dismiss the chi-square test as a formal hypothesis test and rely on other methods to assess fit of the model to the data.

A number of other goodness of fit indices for evaluating fit of the model to the data are available. Goodness of fit indices which are implemented by Mplus include the root mean square residual (RMSR) and the root mean square error of approximation (RMSEA). The residual matrix is also provided.

Fit indices can be classified by whether they are population or sample based. The root mean square error of approximation (RMSEA) is a population based index that is relatively insensitive to sample size. The RMSEA is a measure of the residual variances and covariances, which quantifies the error of approximation of the population data by the model (Loehlin, 1998). Small values of the RMSEA indicate fit, while an RMSEA value of zero would indicate perfect fit. RMSEA values less than .05 indicate very good fit of the factor model to the data, values between 0.05 and 0.08 indicate moderate fit, and those between 0.08 and 0.1 indicate relatively poor fit (Browne & Cudeck, 1993). According to Hu and Bentler (1999), RMSEA values below .06 indicate satisfactory fit of the model to the data.

The root mean square residual (RMR) is the square root of the average square residuals. It summarizes the differences between the observed and expected covariances given the model. Larger values indicate less fit between the model and the data. According to Hu and Bentler (1999) RMR should be below .08 with lower values indicating better fit of the model.

Mplus Version 3.01 using the WLSMV estimator produces the descriptive model fit statistics RMSEA and RMR for categorical data such as the L2/L3 Production Technician test data.

Examination of the residual matrix is another way of assessing the fit of the model to the data. The residual matrix consists of the differences between the observed and hypothesized covariance matrices. Smaller residuals indicate aspects of the data that have been well accounted for by the model while large residuals indicating aspects that are not (Loehlin, 1998). While the previously mentioned measures of fit assess the average fit for the model as a whole, an examination of the residuals can help to identify specific areas of model misfit.

If the model fits the data the next step is to evaluate the model parameters. Most CFA programs impose no constraints to ensure that the estimates have meaningful values (Long, 1983). Therefore, even if the data reveal the fit is acceptable the model parameters must still be evaluated and interpreted. Unreasonable estimates such as negative estimates of variances and/or correlations that exceed plus or minus 1.0 indicate that one of five problems has occurred (Long, 1983).

First, model misspecification may have occurred. Even if the overall fit of the model is acceptable this can still be a problem. Second, violations of the normality assumption for the observed variables may have occurred. This is particularly important for maximum likelihood estimation since there is little evidence of how robust ML is if the assumption of normality is violated (Long, 1983). Third, small samples may be too small to justify the use of the asymptotic properties of the method of estimation which may result in negative estimates of variances (Long, 1983). Fourth, if the model is *nearly* unidentified, the estimation parameters may be unstable. Even if the model can be proven identified, the method of estimation may have a difficult time distinguishing between two or more of the parameters for the sample data (Long, 1983). Fifth, the covariance matrix may have been computed using pairwise deletion of missing data. This is problematic if the covariance or correlation matrices were constructed by using all of the data available for a given pair of variables to compute the covariance or correlation between those two variables. When this occurs each covariance or correlation is based on a different sample, which can lead to a covariance matrix that is inappropriate to use for estimation (Long, 1983).

If the hypothesized model does not fit the data adequately, the model should be respecified (Long, 1983). The results obtained from estimating the rejected model can be used to suggest, additional, hopefully better fitting models. In this process called a specification search, consideration of underlying theory is used in conjunction with an analysis of diagnostic measures. One way to improve the fit of a model is to eliminate nonsignificant parameters as indicated by a z-test. Restricting nonsignificant parameters can improve overall fit by recovering degrees of freedom with little accompanying increase in the χ^2.

Additionally, indicators that were not adequately explained by the model can be identified by an examination of the residual matrix. Large differences may indicate the portion of the model that is misspecified. Latent constructs which were not specified may be added by to the model. However theoretical guidance should always be used to support the addition of one or more latent constructs.

A modification index suggests the expected decrease in the chi-square if a single constraint is relaxed (Long, 1983). By freeing the parameter with the largest modification index the greatest improvement in fit for a model will be obtained. Freeing parameters with modification indices for parameters which are smaller than 3.84 is unlikely to result in a substantial improvement in fit. One parameter at a time should be relaxed only if it makes sense substantively to relax that parameter. Finally, the respecified model should be compared with the hypothesized model to assess improvement in model data fit.

2.7 EXTERNAL VALIDITY EVIDENCE

2.7.1 Relationship Between Performance Assessments and Other Assessment Methods

The relationship between performance assessment measures and other types of assessment methods (e.g. ability and aptitude tests) has been examined previously (Breland & Griswold, 1982; Hattrup & Schmitt, 1990; Hogan, Arneson, & Petersons, 1992). Paper-and-pencil ability tests have been found to predict performance on work sample simulations for high pressure cleaning workers (Hogan, et al., 1992). Hattrup and Schmitt (1990) conducted a criterion-related validation study to assess the validity of four paper-and-pencil aptitude tests and five tests of content taken directly from job tasks in predicting job sample performance of apprentices in eight skilled trades and found observed validities above .40.

2.7.2 Predictive Validity Evidence

The relationship of test scores to variables external to the test provide evidence on whether scores on the test may indeed be interpreted in terms of the construct intended. (AERA et al, 1999) Convergent evidence seeks higher positive correlational evidence with criteria hypothesized to measure the same construct while discriminant evidence seeks lower correlational evidence with measures purportedly of different constructs. Predictive designs involve criteria that will become available some time in the future, while concurrent designs involve criteria that are currently available. In the case of the L2/L3 Production Technician assessment, predictive validity is concerned with the extent to which the MC test forecasts an examinee's future level on the criteria which are the PA exercises.

2.7.3 Linear Regression

Linear regression is a statistical technique that attempts to model the relationship between two variables by fitting a linear equation to observed data (Neter, Kutner, Nachtsheim & Wasserman, 1996). One variable is considered to be the independent or predictor variable, and the other is considered to be a dependent or response variable. While regression models often contain more than one predictor variable, the present study is concerned with only MC test score as a predictor variable due to its central importance as an explanatory variable.

Before attempting to fit a linear model to observed data, it should first be determined whether or not there is a relationship between the variables of interest. Regression analysis only allows one to establish that there is some significant association between the two variables but does not imply that one variable causes the other variables. A scatterplot allows the researcher to investigate the strength of the relationship between two variables. If there appears to be no association between the explanatory and dependent variables then fitting a linear regression model to the data will be of little value. The linear regression model can be stated as follows:

$$Y = a + bX \tag{4}$$

where X is the independent variable and Y is the dependent variable. The slope of the line is b, and a is the intercept. The coefficients a and b are determined by the condition that the sum of the square residuals is as small as possible. The direction of the relationship between variables can be determined by the sign of the B coefficients. If a B coefficient is positive, then the relationship of this variable with the dependent variable is positive; if the B coefficient is negative then the relationship is negative. If the B coefficient is equal to 0 then there is no relationship between the variables.

The regression line expresses the best prediction of the dependent variable (Y), given the independent variables (X). The deviation of a particular point from the regression line (its predicted value) is called the residual value. The smaller the variability of the residual values around the regression line relative to the overall variability, the better the prediction. If there is no relationship between the X and Y variables, then the ratio of the residual variability of the Y variable to the original variance is equal to 1.0. If X and Y are perfectly related, then there is no residual variance and the ratio of variance would be 0.0. In most cases, the ratio would fall somewhere between 0.0 and 1.0. R-square, or the coefficient of determination, is determined by

1.0 minus the ratio of variance (Neter et al., 1996). The R-square value indicates the proportion of the variability accounted for given the variables specified in the model. The R-square value is an indicator of how well the model fits the data where an R-square close to 1.0 indicates that almost all of the variability is accounted for with the variables specified in the model.

Since linear regression assumes that the relationship between variables is linear, a scatterplot of the variables of interest should be examined to assess linearity. Linear regression also assumes that the residuals (predicted minus observed values) are distributed normally. Therefore histograms for the residuals should be inspected for normality.

Regression analysis serves the three, frequently overlapping, purposes of description, control and prediction. A regression analysis can be used to provide external validity evidence by describing and predicting the relationship between the MC test and PA exercises.

2.7.4 Decision Consistency Estimated with Two Tests

Decision consistency refers to the degree to which the same decisions are made from two different sets of measurements (Crocker & Algina, 1986). While Crocker and Algina (1986) refer to the consistency of decisions of two forms of a test or two different administrations of the same test, decision consistency can also be used to evaluate the consistency of decisions based on different types of mastery tests administered to the same group of examinees. Decisions for an examinee are consistent when the results of both tests indicate that an examinee should be classified as passing. Figure 2 shows how pass/fail decision consistency for an MC test and PA test can be determined.

Decisions Based on MC Test

Decisions Based on PA Test	Nonmaster	Master
Nonmaster	$\hat{P}_{00}$	$\hat{P}_{01}$
Master	$\hat{P}_{10}$	$\hat{P}_{11}$

Figure 2. Probabilities of Consistent Classification for Two Tests

The estimated probability of a consistent decision for the MC test and the PA test is calculated by $\hat{P} = \hat{P}_{11} + \hat{P}_{00}$.

A decision consistency analysis provides an additional source of information regarding the relationship between the MC test and PA exercises.

3.0 METHODS

The purpose of this project was to provide validity evidence for the L2/L3 assessment. Specifically, this study provided validity evidence validity evidence for the two alternate equivalent multiple-choice (MC) job knowledge tests and the seven performance assessment (PA) exercises that were developed for the L2/L3 Production Technician at a large consumer products manufacturing company in the southeastern United States. Whereas the MC job knowledge test items were designed to measure what an applicant *knows* about the job, the PA exercises were designed to measure an applicant's ability to *do* or *perform* certain aspects of the job. Based on the review of literature, an examination of the relationship between the two types of items employed for L2/L3 Production Technician selection seemed both timely and appropriate. Additional research was needed to examine the relationship between the MC test items and the PA exercises and to provide evidence to support inferences based upon scores for both assessments. This evidence should demonstrate both the internal structure and the relationship between both types of items.

The sections that follow describe the research questions and the methods, including the procedures for data collection and analysis, to address each research question.

The research questions that were investigated in this study:

1. Based on a content analysis of the assessments, what knowledge, skills, and abilities are measured by the MC job knowledge test and the PA exercises?

1a. Based on a content analysis of the assessments, to what extent do the PA exercises measure the same knowledge, skills, and abilities as the MC job knowledge test items?

1b. Based on a content analysis of the assessments, what additional knowledge, skills, or abilities are being assessed by the PA exercises beyond what is measured by the MC job knowledge test items?

1c. Based on the results of the content analysis of the assessments, which subtests of the MC job knowledge tests are most related to each of the seven PA exercises?

2. What are the item and test properties of the MC job knowledge test items and the PA exercises?

   2a. To what degree are item and test properties similar for both Form A and Form B?

3. What is the factor structure of both Form A and Form B with the PA exercises included?

4. What is the relationship between scores on the MC job knowledge tests and total score on the PA exercises?

   4a. Based on a regression analysis, what is the nature of the relationship between the predictor variable MC test score and the dependent variable PA test score?

   4b. What is the decision consistency regarding those who score above and below the cut score on the MC test and the PA exercises?

5. What is the relationship between each subtest of the MC job knowledge test and each of the seven PA exercises?

3.1 OVERVIEW OF L2/L3 PRODUCTION TECHNICIAN ASSESSMENT DEVELOPMENT

3.1.1 Problem and Setting

A large consumer products manufacturing company in the southeastern United States planned to close an older facility and move its production and maintenance workers to a new facility located in the same city. Because the new facility contained more advanced equipment and technology compared to the older facility, job duties at the new facility were also found to be more complex. As a consequence, management wished to ensure that workers at the new facility possessed the necessary knowledge, skills, and abilities to perform the duties of the newer, more advanced jobs. At management's request, a team of Industrial/Organizational Psychologists and other consultants was assembled to develop an assessment procedure which would be used to select those current workers who were qualified to perform the jobs in the new facility.

As referred to in the present study, the original job at the older facility was L1 Team Member. Management sought to give qualified L1 workers at the old facility a chance to move into jobs at the new facility. Personnel seeking employment at the new facility would need to be qualified to perform the work of the job, referred to in the present study, as L2 Production Technician and would have the opportunity to move up to a higher job, referred to in the present study, as L3 Production Technician. The Job Summaries, Work Performed, and Consequences of Errors for L1 Team Member (Ramsay, 1999), L2 Production Technician (Ramsay, 2000a) and L3 Production Technician (Ramsay, 2000b) are shown in the original validation reports.

3.1.2 L1, L2, L3 Job Analysis

In October of 1999 a team of Industrial Psychologists conducted job analyses for the jobs of L1 Team Member, L2 Production Technician, and L3 Production Technician at the company's facility in the southeastern United States.

In order to ensure that assessment would reflect the knowledge and skills required on the job, the content-related validation model was employed. The job analysis activities were designed to identify the important work behaviors and necessary knowledge, skills, and abilities (KSAs) required for learning and performing the job. Four job analysis methods were used to study the jobs. First, a review of company documents (e.g., job descriptions, training programs, standard operating procedures, and employee handbook) was conducted. Second, job analysts directly observed the jobs being performed. Third, the consultants conducted group meetings with job experts (e.g., managers, supervisors, engineers, trainers). Job incumbents were not used as job experts because their personal interests may have conflicted with company interests. Fourth, the Position Analysis Questionnaire (PAQ), a standardized job analysis instrument designed for hourly production jobs, was used. The PAQ report helped to document the similarities and differences among the L1, L2, and L3 positions (Seberhagen, 1999a,1999b,1999c). Results of the job analyses are reported in the original validation reports for L1 Team Member, (Ramsay, 1999), L2 Production Technician (Ramsay, 2000a) and L3 Production Technician (Ramsay, 2000b).

L2 and L3 job experts were selected according to three criteria: (a) they could *not* be current incumbents nor be related to current incumbents, (b) they could be technicians, engineers, or managers, (c) they should be as diverse as possible in race, gender, and national origin (while excluding incumbents or relatives of incumbents).

3.1.3 Selection of Assessment Content

In consultant led group meetings, the job experts were asked to generate in a list of KSAs they felt were important for the positions of L1 Team Member, L2 Production Technician and L3 Production Technician. Items found to be statistically significant on the *Job Activity Checklist (JAC)* (Ramsay, 1970) were combined with the job expert generated lists of KSAs to develop the final sets of KSAs for each of the three jobs shown in Appendix A. A copy of the *Job Activity Checklist (JAC)* is shown in Appendix B. A summary of the *JAC* methodology and results is shown in Appendix C (Ramsay, 1999, 2000a, 2000b).

The results of the job analyses revealed that the L1 Team Member differed significantly from L2 Production Technician and the L3 Production Technician. As a result, the KSAs for the L2 Production Technician and for the L3 Production Technician were the basis for the development of assessments for the two jobs.

After consultation with both management and job experts, it was decided that a selection procedure consisting of both MC paper-and pencil and hands on PA elements would be the fairest and most objective means of evaluation.

Two 125-item alternate tests using an MC format were initially developed (Form A and Form B) to assess L2/L3 Production Technician knowledge. Additionally, 7 PA exercises were developed to evaluate skills in the following areas: Plate Alignment, Cylinder Alignment, Automatic Sequence, Pneumatic System (Vacuum), Pneumatic System (Cylinder Speed), Component Connection, and Electrical Circuit Test.

3.2 DESCRIPTION OF L2/L3 PRODUCTION TECHNICIAN ASSESSMENTS

3.2.1 MC Job Knowledge Test

A detailed Job Task Analysis was conducted to generate a list of Knowledge and Skill areas suitable for use in developing a paper and pencil multiple-choice test for L2 Production Technician and L3 Production Technician. Job experts were asked as group to edit this list and then rank each Knowledge and Skill Areas in terms of its importance. These rankings were then averaged across the raters. Next job experts were asked to estimate the percentage of items they would prefer to be included in each Knowledge and Skill area based on the group's Average Ranked Importance. The Average Percent of Items was calculated by averaging the estimated percent of items across raters. Finally, the Average Percent of Items for each Knowledge and Skill Area was multiplied by the number of items planned for the test (120 items). Table 1 shows the Knowledge and Skill Areas, the Average Ranked Importance, Average Percent of Items, and Number of Questions for each knowledge and skill area for the planned L2/L3 Production Technician Test (Ramsay, 2000c).

Table 1. L2/L3 Production Technician Knowledge and Skill Areas

Average Ranked Importance	Average % of Items	Estimated No. of Questions	Knowledge and Skill Area
2	20.8	25	A. Mechanical 1. Troubleshoot 2. Repair/replace 3. Principles 4. Operations 5. Flow paths 6. Alignment 7. Gear boxes 8. Conveyors 9. Pumps 10. Valves 11. Assembly drawings
4	16.7	20	B. Electrical 1. Troubleshoot 2. Repair/replace 3. Servos 4. Switches 5. Heaters 6. Motors 7. Blowers/fans 8. AC/DC circuits 9. PLC systems 10. Electrical drawings
7	5.8	10	C. Hydraulics (Fluid Flow) 1. Troubleshoot 2. Repair/replace 3. Pumps 4. Accumulators 5. Control valves 6. Hoses 7. Hydraulic prints
6	6.7	10	D. Hand/Measuring Tools 1. Wrenches 2. Multimeters 3. Voltmeters 4. Calipers 5. Gauges
5	8.3	10	E. Math/Statistics 1. X/Y axis 2. Percentages 3. Decimals 4. Fractions 5. Averages
1	24.2	29	F. Problem Solving 1. Cause & effect 2. Interpret data 3. Resolve issues 4. Bar/line graphs 5. Pareto & pie charts 6. Flow diagrams
3	17.5	21	G. Pneumatics 1. Troubleshoot 2. Repair/replace 3. Air cylinders 4. Air filters 5. Hoses 6. Nozzles 7. Valves 8. Vacuum systems 9. Fittings/couplings
		125	Safety was found to be pervasive

Note. From *Content validation report: Assessment selection & development [L2&L3 (Production Technician)]* by R.T. Ramsay, 2000, p. 11.

3.2.2 Selection and Development of Test Items

Working in pairs, job experts were given test questions from Ramsay Corporation's database and were asked to select questions, answer them, and for each item provide a one- or two-word description for each of two alternate forms. New items were written if suitable questions could not be found in Ramsay Corporation's database. Job experts were given instruction on making the two test forms as similar as possible. The tests were then edited and composed by Ramsay Corporation.

3.2.3 Selection Procedures and Their Content

Table 2 shows the final Knowledge and Skill Areas and the final numbers of items for the test.

Table 2. Items in Each Test Area for L2/L3 Production Technician Form A & Form B

Area	Actual No. of Items
Mechanical	(25)* 20
Electrical	(20)* 30
Hydraulics	10
Hand/Measuring Tools	(10)* 8
Math/Statistics	10
Problem Solving	(29)* 26
Pneumatics	21
Total	125

* Changed by consensus of job experts at the time of Angoff and Job Relatedness ratings.

Note. Adapted from *Content validation report: Assessment selection & development [L2&L3 (Production Technician)]* by R.T. Ramsay, 2000, p. 13.

3.2.4 Development of Performance Assessment Measures

Job experts were given the knowledge, skills, and abilities and asked to suggest exercises which could be used to assess skills of assesses in a fair, objective, and valid manner in a large-scale setting. Eight different exercises were initially suggested. After considerable review and comment, seven of the eight were chosen by job experts to be the final performance assessment exercises. These final 7 performance assessment exercises are Plate Alignment, Cylinder Alignment, Automatic Sequence, Pneumatic System (Vacuum), Pneumatic System (Cylinder Speed), Component Connection, and Electrical Circuit Test.

Table 3 shows a brief description of the performance assessment measures.

Table 3. Description of Performance Measures

Name	Station	Description
Plate Alignment	1	• A person will set an open vertical gap between two plates using a jig. A 4-point adjustment is required. • The bottom plate must be aligned on both axes of the horizontal plane using threaded adjustment devices. Alignment pins will fit into bushings when the top plate is lowered.
Cylinder Alignment	2	• A person will align the body and rod of a pneumatic cylinder to a parallel surface using a jig. • The cylinder stroke or cushion will be adjusted using a jig. • All adjustments will be driven by the improper operation of a proximity switch.
Automatic Sequence	3	• A person will evaluate the operation of an automated sequence. • The system will use photo sensors, proximity switches, vacuum switches and a pneumatic cylinder to pick up a flat object. • Various sensor and/or switch faults will identify the (one) faulty component. • Simple automation will be used but PLC troubleshooting is not required.
Pneumatic System (Vacuum)	4	• A person will evaluate a pneumatic circuit (Vacuum side). • Vacuum setting, vacuum cup integrity, air supply to the vacuum generator, along with pressure and vacuum switch adjustment are possible. Only one variable will be adjusted at a station.
Pneumatic System (Cylinder Speed)	5	• A person will evaluate and adjust the operation of a pneumatic cylinder. • Supply pressure, directional control valve operation, metering valve settings, hose and filter flow are causes that must be identified and corrected. Only one variable will be adjusted at a station.
Component Connection	6	• A person will install a directional control valve and a pneumatic cylinder into a system including the attachment of air lines. • Component and system integrity (no leaks) will be checked. • Cylinder stroke with solenoid operation will be evaluated: If Solenoid A energizes, then the cylinder will extend.
Electrical Circuit Test	7	• A person will troubleshoot the electric circuit shown in the diagram to determine the location of an open in the circuit. • The opens in the circuit will be simulated by toggle switches placed in series with the output of each of the components in the circuit.

Note. Adapted from *Content validation report: Assessment selection & development [L2&L3 (Production Technician)]* by R.T. Ramsay, 2000, p.15-16.

3.3 MODIFIED ANGOFF SESSIONS

3.3.1 MC Job Knowledge Test

After the questions were composed and printed, the multiple-choice job knowledge tests were presented to the job experts for review and determination of cutting scores. A modification of Angoff's method (Livingston and Zieky, 1982) was used to determine the cutting score. The job experts took turns answering each question and indicating the percent of qualified persons who would get the item correct. They were then told the percent passing in the largest group to whom the question was given. Their responses were averaged and became the cutting score, which turned out to be 100 of 125 items.

3.3.2 PA Measures

The job experts were given a description of the performance assessment exercises. They then judged what percent of qualified L2 or L3 employees would get that exercise correct. These judgments were averaged and multiplied times the number of assessments. The resulting cutting score turned out to be 6 of 7.

3.3.3 Job Relatedness of the Job Knowledge MC Test

As part of the modified Angoff procedure, job experts were asked to indicate the job relatedness of each test item according to the following key: (a) 5 = Very High, (b) 4 = High, (c) 3 = Medium, (d) 2 = Low, (e) 1 = Very Low, (f) 0 = None. The average job relatedness for Form A was 4.0 and for Form B was 4.1, respectively. None of the items had an average job relatedness below 2.5.

3.3.4 Job Relatedness of the PA Measures

As part of the modified Angoff procedure, job experts were asked to indicate the job relatedness of each test item according to the same key used for the MC test. The average job relatedness for the seven PA measures is shown in Table 4.

Table 4. Average Job Relatedness for PA Measures

Exercise	**J.R. Mean**
1	4.8
2	4.7
3	4.5
4	4.3
5	4.5
6	4.3
7	4.8
Total Mean	4.56

Note. Adapted from *Content validation report: Assessment selection & development [L2&L3 (Production Technician)]* by R.T. Ramsay, 2000, Appendix G, p.4.

The data in Table 4 indicate that the job experts rated the PA very highly in job relatedness.

3.4 TEST ADMINISTRATION AND SCORING

In the vast majority of validation studies conducted by Ramsay Corporation in which PA exercises are developed along with a MC test, examinees take the PA portion only if they have passed the MC test. The test administration for L2/L3 Production Technician was unique in that the tests were not scored until all applicants took both the MC portion and the PA portion at the same test administration session. However, both the MC portion and the PA exercises were treated as two separate tests with regards to the application of the cut scores. In other words, examinees had to score above both individual cut scores to be considered qualified.

The MC portion of the test had no time limit but examinees were told that they should not need more than two hours to complete the test.

Each of the seven PA exercises was considered as a separate station. Each exercise had a ten minute time limit. There was one administrator per station and the administrator was the sole scorer of the exercise.

3.5 DESCRIPTION OF SAMPLE

The study sample included examinees who took either Form A or Form B of the MC job knowledge test and the PA exercises between March of 2000 and April of 2002. All examinees were required to take both the MC job knowledge test and the PA exercises during the same test administration session.

3.5.1 Form A

For Form A, there were 3 examinees that took the job knowledge test but did not take the PA exercises. The final sample for the 432 applicants who took both Form A and the PA exercises between March of 2000 and April of 2002 was comprised of 382 males and 50 females. There were 309 Whites, 62 Blacks, 34 Asians, and 27 Hispanics.

3.5.2 Form B

For Form B, there were 2 examinees that took the job knowledge test but did not take the PA exercises. The final sample for the 324 applicants who took both Form A and the PA exercises between March of 2000 and April of 2002 was comprised of 283 males and 41 females. There were 211 Whites, 62 Blacks, 35 Asians, and 16 Hispanics.

A summary of the research questions and methods that were used to answer each question appears in the following sections.

3.6 RESEARCH QUESTION 1 - BASED ON A CONTENT ANALYSIS OF THE ASSESSMENTS, WHAT KNOWLEDGE, SKILLS, AND ABILITIES ARE MEASURED BY THE MC JOB KNOWLEDGE TEST AND PA EXERCISES?

Two subject-matter experts evaluated the content of Form A and the PA exercises. The content analysis was conducted only on Form A since both forms Form A and Form B were constructed to be alternate-equivalent tests. In accordance with the recommendation by Rovinelli and Hambleton (1976) to utilize a simple technique that would not be tedious or time consuming to the content specialist, a version of the semantic differential procedure was employed. The subject-matter experts evaluated each MC item in terms of its content and its relationship to each of the seven PA exercises according to a 4-point Likert scale where: (a) 0 = no relationship, (b) 1 = small relationship, (c) 2 = moderate relationship, and (d) 3 = strong relationship.

3.6.1 1a. Based on a content analysis of the assessments, to what extent do the PA exercises measure the same knowledge, skills, and abilities as the MC job knowledge test items?

As suggested by Hambleton (1984), after each of the MC items was rated by the subject matter experts, the ratings were averaged over the number of subject matter experts and compiled to determine the mean relevance rating of each of the seven PA exercises to the total MC test. Although each individual MC item and its relevance to each of the seven PA exercises was calculated, the results compiled and analyzed by (a) total MC test with total on PA exercises, (b) total MC test with each PA exercise, and (c) each MC subtest with each PA exercise, were of primary interest to the present study. The results were evaluated and compared to the results in research question 4.

3.6.2 1b. Based on a content analysis of the assessments, what additional knowledge, skills or abilities are being assessed by the PA exercises beyond what is measured by the MC job knowledge test items?

As part of the content analysis, the two subject-matter experts were asked to identify for each of the PA exercises if any additional knowledge, skills, or abilities are being assessed beyond what is measured by the MC test items. The results for the two job experts were compared and evaluated. Additionally, the subject-matter experts were asked to rate the importance of the seven PA exercises in terms of their importance to the job of L2/L3 Production Technician according to the following scale: (a) 0 = not important (b) 1 = small importance, (c) 2 = moderate importance, and (d) 3 = great importance.

3.6.3 1c. Based on the results of the content analysis of the assessments, which subtests of the MC job knowledge tests are most related to each of the seven PA exercises?

The results of the subject-matter experts' content analysis and ratings were averaged and compiled for each of the MC test sections with each of seven PA exercises. The results were evaluated and compared to the results in research question 5.

3.7 RESEARCH QUESTION 2 - WHAT ARE THE ITEMS AND TEST PROPERTIES OF THE MC JOB KNOWLEDGE TEST FORMS AND THE PA EXERCISES?

The software program ITEMAN was used to conduct classical item analyses for the MC test and the seven exercise PA test. Item difficulty, and point biserial correlations were evaluated for each item and exercise. Mean, standard deviation, standard error of measurement, and coefficient alpha were calculated and evaluated. Additionally, skewness, kurtosis, and frequency distributions in the form of histograms were compared and evaluated.

3.7.1 2a. To what degree are item and test properties similar for both Form A and Form B?

If *Lord's equity property* of equating (Lord, 1980) holds for the two forms of the L2/L3 Production Technician Test then it does not matter whether an examinee takes Form A or Form B. This property implies that examinees with a given true score would have identical observed score means, standard deviations, and distributional shapes of converted scores on Form A and on Form B (Kolen & Brennan, 1995). The implication of identical standard deviations means that the standard error of measurement at any true score will be equal on both forms (Kolen & Brennan, 1995). In reality, however, perfectly identical forms typically cannot be constructed.

A comparison of classical item analyses for Form A and Form B was conducted. It was hypothesized that results of the item analyses will be similar for both forms since Form A and Form B were designed to be alternate equivalent versions of the same test and, as such, both forms share the exact same test blueprint.

3.8 RESEARCH QUESTION 3 - WHAT IS THE FACTOR STRUCTURE OF BOTH FORM A AND FORM B WITH THE PA EXERCISES INCLUDED?

Additional research was needed to examine the relationship among the MC test items and the PA exercises and to provide evidence to support inferences based upon scores for the assessments. This evidence should demonstrate both the internal structure and the relationship among both types of assessment items.

Factor analysis assumes that the observed variables are linear combinations of some unobservable, underlying factor (Kim & Mueller, 1978). Factor analysis studies can be either exploratory or confirmatory. In exploratory factor analysis (EFA) the objective is to try to find a factor structure that could account for the intercorrelations of an observed set of variables. Since a compelling theory of the underlying structure of the variables was not readily apparent in the present study, an EFA using Mplus version 3.11 was conducted first to investigate the underlying patterns of the data.

The factor extraction method, weighted least squares means and variances adjusted (WLSMV) was used since the L2/L3 Production Technician test data are categorical.

The number of factors to retain for rotation was determined by an analysis of the graph of the eigenvalues or scree plot. The scree test involves visually inspecting the graph of the eigenvalues, and looking for the natural bend or break point in the data where the curve flattens out. The number of datapoints above the break, not including the point at which the break occurs is usually the number of factors to retain.

The initial factor loadings were rotated in an attempt to find the simplest and most easily interpreted factor structure. The oblique rotation method, Promax, was used since this method allows the factors to correlate and it was expected that there would be some correlation among factors. Finally, an attempt was made to interpret or explain the factor structure.

Confirmatory factor analysis (CFA) is a means for grouping items into content or process categories. It can be used to verify the reasoning that goes into test specifications, providing empirical evidence for the content or process categories of a test. In the confirmatory factor model, the researcher must determine in advance of analysis which constraints to impose including (a) which pairs of common factors are correlated, (b) which observed variables are affected by which common factors, (c) which observed variable are affected by a unique factor, and (d) which pairs of unique factors are correlated (Long, 1983). CFA also allows that statistical tests be performed to determine if the sample data are consistent with the imposed constraints (Long, 1983).

The computer program Mplus Version 3.11 was used to conduct a confirmatory factor analysis on the MC items combined with the PA exercises. Three different competing models were evaluated. Model 1 hypothesized that the one-factor model would fit the MC test combined with the PA exercises since both portions of the test were designed to measure overall *applied mechanical knowledge*. Model 2 hypothesized that there are two different dimensions underlying the test scores. The two dimensions consist of separate dimensions for the MC portion and the PA exercises. Model 3 hypothesized that there are eight different dimensions underlying the test scores. The eight dimensions consist of the seven content areas of the MC portion (Mechanical, Electrical, Hydraulics, Hand/Measuring Tools, Math/Statistics, Problem Solving, and Pneumatics) plus one additional dimension for the PA exercises.

The CFA models were evaluated for both Form A and Form B. It was hypothesized the CFA results for both forms would be the same due to the fact that Form A and Form B were designed to be alternate equivalent versions of the same test and as such, both forms share the exact same test blueprint.

3.9 RESEARCH QUESTION 4 - WHAT IS THE RELATIONSHIP BETWEEN SCORES ON THE MC JOB KNOWLEDGE TESTS AND TOTAL SCORE ON THE PA EXERCISES?

The Pearson product moment correlation coefficient is a measure of the degree of linear relationship between two sets of observations. The raw-score formula for the correlation coefficient is:

$$\rho_{XY} = \frac{\Sigma(X - \mu_X)(Y - \mu_Y)}{N\sigma_X\sigma_Y} \qquad (9)$$

where, X is a raw score on variable X, Y is a raw score on variable Y, μ_X is the mean of variable X, μ_Y is the mean of variable Y, N is the number of persons, σ_X is the standard deviation of variable X scores, and σ_Y is the standard deviation of variable Y scores (Crocker & Algina, 1986). Values of the correlation coefficient can range from -1.00 to 1.00 with the sign of the number indicating the positive or negative direction of the relationship. Values of ρ_{XY} that are around .00 indicate little or no relationship between the variables X and Y (Crocker & Algina, 1986).

Correlations between total score on the MC test items and total score on the PA exercises were calculated and evaluated. Similarly, correlations between each subtest of the job knowledge test and the total score on the PA exercises were calculated and evaluated. It was hypothesized that both Form A and Form B would reveal strong positive correlations with total score on the PA exercises as both types of assessments were designed to measure essentially the same skills albeit in different formats. Furthermore, past research has shown significant positive correlations between various assessment types and PA assessments (Breland & Griswold, 1982; Hattrup & Schmitt, 1990, Hogan, Arneson, & Petersons, 1992).

3.9.1 4a. Based on a regression analysis, what is the relationship between the predictor variable MC test score and the dependent variable PA test score?

Linear regression is a statistical technique that attempts to model the relationship between two variables by fitting a linear equation to observed data. One variable is considered to be the independent or explanatory variable, and the other is considered to be a dependent variable. A regression analysis was conducted using MC test scores as the independent variable and the PA test as the dependent variable. In order to examine linearity, a scatterplot of MC test scores (the independent variable) with PA test scores (the dependent variable) along with the computed regression line was plotted. SPSS 13.0 for Windows also calculates the F statistic which tests the hypothesis that there is no linear relationship between X and Y.

The R-square value was calculated and evaluated. The R-square value indicates the amount of the variability accounted for given the variables specified in the model. The R-square value is an indicator of how well the model fits the data where an R-square close to 1.0 indicates that almost all of the variability is accounted for with the variables specified in the model.

3.9.2 4b. What is the decision consistency regarding those who score above and below the cut score on the MC test and the PA exercises?

Decision consistency refers to the degree to which the same decisions are made from two different sets of measurements (Crocker & Algina, 1986). Decisions for an examinee are consistent when the results of both the MC test and the PA test indicate that an examinee should be classified as passing. The percentages of consistent and inconsistent classifications from the MC test and the PA test were calculated and evaluated. A consistent decision is calculated by summing the number of examinees who passed both the MC test and the PA test with the number of examinees who failed both tests.

3.10 RESEARCH QUESTION 5 - WHAT IS THE RELATIONSHIP BETWEEN EACH SUBTEST OF THE MC JOB KNOWLEDGE TEST AND EACH OF THE SEVEN PA EXERCISES?

Correlations between each subtest of the MC test items and scores on each of the PA exercises were calculated and evaluated. It was hypothesized that subtests on both Form A and Form B would show moderate positive correlations with the each of the PA exercises as they are all measuring some aspect of a construct labeled *applied mechanical knowledge*, albeit in different formats.

4.0 RESULTS

4.1 RESEARCH QUESTION 1 - BASED ON A CONTENT ANALYSIS OF THE ASSESSMENTS, WHAT KNOWLEDGE, SKILLS, AND ABILITIES ARE MEASURED BY THE MC JOB KNOWLEDGE TEST AND PA EXERCISES?

Two subject-matter experts evaluated the content of Form A and the seven PA exercises. The content analysis was conducted only on Form A since both forms Form A and Form B were constructed to be alternate-equivalent tests.

4.1.1 Subject-matter Expert Qualifications

Subject-matter expert Number One was an Account Director for a training organization and possessed eleven years of manufacturing or maintenance work experience and seven years of experience supervising or training manufacturing or maintenance workers. Subject-matter expert Number Two was a Project Manager for a training organization and possessed with twenty years of manufacturing or maintenance work experience and six years of experience supervising or training manufacturing or maintenance workers. Subject-matter expert Number One indicated that he held a B.S. degree in Management and completed a two year Navy Electronics and Nuclear program. Subject-matter expert Number Two indicated that he held a B.S. degree in Nuclear Engineering Technology and also completed the U.S. Navy Electronics and Nuclear program.

4.1.2 1a. Based on a content analysis of the assessments, to what extent do the PA exercises measure the same knowledge, skills, and abilities as the MC job knowledge test items?

4.1.2.1 Mean Relevance Ratings The subject-matter experts evaluated each MC item in terms of its content and its relationship to each of the seven PA exercises according to a 4-point Likert scale where: (a) 0 = no relationship, (b) 1 = small relationship, (c) 2 = moderate relationship, and (d) 3 = strong relationship. The subject-matter expert survey with instructions is shown in Appendix D.

As suggested by Hambleton (1984), after each of the MC items was rated by the subject matter experts, the ratings were averaged over the number of subject matter experts and compiled to determine the mean relevance rating of each of the seven PA exercises to the total MC test. The results compiled and analyzed by (a) total MC test with total on PA exercises, and (b) total MC test with each PA exercise. The mean relevance ratings of the total MC test with PA1 through PA7 were 0.16, 0.24, 0.52, 0.46, 0.47, 0.44 and 0.21 respectively. The mean relevance rating of the total MC test with the total PA test was 0.36. PA3 (Automatic Sequence) had the highest mean relevance rating 0.52 with the MC test, while PA1 (Plate Alignment) had the lowest mean relevance rating 0.16 with the MC test.

4.1.2.2 Generalizability Theory SPSS 13.0 was used to generate variance component estimates required for the calculation of the G-coefficient for a $p \times r$ design for each of the seven PA exercises. The G-coefficient, which is similar to the reliability coefficient in classical test theory, is the proportion of expected observed-score variance that is universe-score variance (Shavelson & Webb, 1991). In a typical $p \times r$ design, persons (p) are the targets of measurement, and rater (r) is treated as a random facet. In the present study however, MC items (mc_item) are the targets of measurement and the intent is to generalize the measurement across the random facet: raters (rater). The variance estimates, percent of total variance and the G-coefficients (ρ^2) for the seven *mc_item* $\times$ *rater* studies are shown in Table 5.

Table 5. Variance Estimates and G-Coefficients for MC Item x Rater

PA	Component	Estimate	Percent of Total Variance	ρ^2
1	Var(mc_item)	0.064	19%	
	Var(rater)	0.029	9%	
	Var(Error)	0.239	72%	.323
2	Var(mc_item)	0.111	23%	
	Var(rater)	0.057	12%	
	Var(Error)	0.307	65%	.379
3	Var(mc_item)	0.207	31%	
	Var(rater)	0.083	12%	
	Var(Error)	0.381	57%	.472
4	Var(mc_item)	0.234	35%	
	Var(rater)	0.029	4%	
	Var(Error)	0.411	61%	.515
5	Var(mc_item)	0.257	30%	
	Var(rater)	0.177	21%	
	Var(Error)	0.419	49%	.463
6	Var(mc_item)	0.121	16%	
	Var(rater)	0.127	16%	
	Var(Error)	0.529	68%	.269
7	Var(mc_item)	0.123	30%	
	Var(Rater)	0.029	7%	
	Var(Error)	0.255	63%	.464

Dependent Variable: Rating
Method: ANOVA (Type III Sum of Squares)

For all seven of the studies, the largest source of variance was for the highest order interaction mc_item $\times$ rater (as well as residual error) which is labeled Var(Error). The Var(Error) ranged from 49 to 72 percent. The smallest source of variance for all of the seven studies was for rater which ranged from 4 to 21 percent of the total variance.

The G-coefficient values which range from 0.269 to 0.515 are rather low and are likely due to the small (2) number of raters used in the present study as well as the somewhat restricted range of rater responses. In fact, one subject-matter offered that as he looked at each individual question and how it related to each PA exercise, he often saw very little relationship or overlap between the two types of items (at least at the individual item level). However, the subject-matter expert considered the MC test as a whole a good predictor of performance on the seven PA exercises.

4.1.3 1b. Based on a content analysis of the assessments, what additional knowledge, skills or abilities are being assessed by the PA exercises beyond what is measured by the MC job knowledge test items?

As part of the content analysis, the two subject-matter experts were asked to identify for each of the PA exercises if any additional knowledge, skills, or abilities are being assessed beyond what is measured by the MC test items. Additionally, the subject-matter experts were asked to rate the importance of the seven PA exercises in terms of their importance to the job of L2/L3 Production Technician according to the following scale: (a) 0 = not important (b) 1 = small importance, (c) 2 = moderate importance, and (d) 3 = great importance. Typed versions of the subject-matter expert's survey responses are shown in Figures 3 and 4.

The responses from the two subject-matter experts indicated that the PA exercises measure a more applied understanding of the subject matter than the MC test. Both subject-matter experts identified specific applications of job knowledge, and demonstrations of proper skills and techniques that they felt were beyond what was measured by the MC test. This was expected since the MC tests were designed to measure an examinee's knowledge of job-related information, whereas the PA exercises were designed to demonstrate an examinee's ability to apply relevant job knowledge.

Rater Code	A1
Date Rated	2/17/2005

Name	Importance to the Job (Circle One)	Please list or describe any additional Knowledge, Skills, and Abilities that are being assessed by these exercises that are beyond those that are being measured by the multiple-choice test items.
1. Plate Alignment	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Skill in the application of alignment principles and techniques
2. Cylinder Alignment	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Skill in the application of alignment principles and techniques Knowledge of location, function, and operation of pneumatic air cylinders Knowledge of location, function, and operation of proximity switches Skill in the application of techniques for troubleshooting and replacing proximity switches
3. Automatic Sequence	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Knowledge of location, function, operation of photosensors Knowledge of location, function, and operation of proximity switches Knowledge of location, function, operation of solenoid valves
4. Pneumatic System (Vacuum)	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Knowledge and skill in the application of pneumatic principles Skill in the reading and interpretation of the pneumatic system flowpath Knowledge and skill in the application of vacuum system principles Knowledge of vacuum system flowpath Knowledge of location, function, and operation of vacuum system piping and connectors
5. Pneumatic System (Cylinder Speed)	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Knowledge and skill in the application of pneumatic principles Knowledge of location, function, and operation of pneumatic hoses Knowledge of location, function, and operation of pneumatic air cylinders
6. Component Connection	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Knowledge and skill in the application of pneumatic principles Knowledge of location, function, and operation of pneumatic hoses Knowledge of location, function, and operation of pneumatic air cylinders
7. Electrical Circuit Test	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Knowledge and skill in the application of electrical principles Knowledge of location, function, and operation of electric circuits Skill in the application of techniques for troubleshooting and replacing electric circuits Skill in the reading and interpretation of wiring diagrams

Figure 3. Job Expert Number One's Survey

Rater Code	A2
Date Rated	4/13/2005

Name	Importance to the Job (Circle One)	Please list or describe any additional Knowledge, Skills, and Abilities that are being assessed by these exercises that are beyond those that are being measured by the multiple-choice test items.
1. Plate Alignment	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Although torque specs were not given in the procedure, one of the steps stated not to "overtorque" the alignment screws. Knowledge and skills on proper torqueing may be assessed with this exercise.
2. Cylinder Alignment	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Can't think of any.
3. Automatic Sequence	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Can't think of any.
4. Pneumatic System (Vacuum)	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Pneumatic regulator adjustment to obtain proper vacuum. How to read vacuum gages. Component identification. Understanding "flow" through diagram or shcematic.
5. Pneumatic System (Cylinder Speed)	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Purpose of limit switches.
6. Component Connection	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Pneumatic system safety. Understanding flow through a solenoid control valve.
7. Electrical Circuit Test	(3) great importance **2** = moderate importance **1** = small importance **0** = not important	Basic troubleshooting stem-symptom recognition. Understanding/reading an electrical schematic diagram. (Question #47 covered pc ladder logic, which is not exactly the same.)

Figure 4. Job Experts Number Two's Survey

Since both subject-matter-experts rated every PA exercise a “3 = great importance” in terms of their importance to the job of L2/L3 Production Technician no additional statistical analyses were done.

4.1.4 1c. Based on the results of the content analysis of the assessments, which subtests of the MC job knowledge tests are most related to each of the seven PA exercises?

The subject-matter experts evaluated each MC item in terms of its content and its relationship to each of the seven PA exercises according to a 4-point Likert scale where: (a) 0 = no relationship, (b) 1 = small relationship, (c) 2 = moderate relationship, and (d) 3 = strong relationship. The results of the subject-matter experts’ content analysis and ratings were averaged and compiled for each of the MC test sections with each of seven PA exercises. The results are shown in Table 6.

Table 6. Subject-matter Expert Content Analysis Rating

	PA1	PA2	PA3	PA4	PA5	PA6	PA7	Mean
Mechanical	0.65	0.68	0.45	0.75	0.70	0.70	0.10	0.58
Electrical	0.00	0.05	0.30	0.10	0.10	0.48	0.72	0.25
Hydraulics	0.00	0.25	0.30	0.25	0.50	0.25	0.05	0.23
Hand Tools	0.56	0.69	0.06	0.19	0.25	0.06	0.00	0.26
Math/Stats	0.15	0.10	0.60	0.00	0.25	0.00	0.00	0.16
Problem Solving	0.00	0.00	0.92	0.42	0.13	0.29	0.10	0.27
Pneumatics	0.02	0.26	0.64	1.14	1.36	0.76	0.00	0.60

PA1 (Plate Alignment) had the highest mean relevance rating (0.65) with the Mechanical subsection of the MC test. PA2 (Cylinder Alignment) had the highest mean relevance rating with the Mechanical (0.68) and the Hand Tools (0.69) subsections of the MC test. PA3 (Automatic Sequence) had the highest mean relevance rating (0.92) with the Problem Solving

subsection of the MC test. PA4 (Pneumatic System – Vacuum) had the highest mean relevance rating (1.14) with the Pneumatics subsection of the MC test. PA5 (Pneumatic System – Cylinder Speed) had the highest mean relevance rating (1.36) with the Pneumatics subsection of the MC test. PA6 (Component Connection) had the highest mean relevance rating (0.76) with the Pneumatics subsection of the MC test. PA7 (Electrical Circuit Test) had the highest mean relevance rating with the Electrical subsection of the MC test.

The Mechanical and Pneumatics subsections of the MC test had the highest overall mean relevance ratings (0.58 and 0.60 respectively) with the seven PA exercises. The Math/Statistics subsection of the MC test had the lowest overall mean relevance ratings of 0.16 with the seven PA exercises.

Although the mean relevance ratings are low overall, the results appear to confirm what the test developers intended to measure with the PA exercises. That is, there are stronger relationships between MC subsections and PA exercises that feature the same knowledge or skill area (e.g., PA7 (Electrical Circuit Test) had the highest mean relevance rating with the Electrical subsection of the MC test).

4.2 RESEARCH QUESTION 2 - WHAT ARE THE ITEMS AND TEST PROPERTIES OF THE MC JOB KNOWLEDGE TEST FORMS AND THE PA EXERCISES?

4.2.1 Test Properties

The software program ITEMAN was used to conduct classical item analyses for the two forms of the MC test and the seven exercise PA test. Mean, standard deviation, standard error of measurement, and coefficient alpha were calculated for both MC test forms and the PA exercises. Additionally, skewness, kurtosis, and frequency distributions in the form of histograms were produced. The overall descriptive statistics for Form A, Form B, and the seven PA exercises are shown in Table 7.

Table 7. Item Analysis Summary

	Form A	Form B	PA Exercises
N of Items	125	125	7
N of Examinees	432	324	756
Mean	95.637	94.512	4.089
Variance	191.981	222.238	3.226
Std. Dev.	13.856	14.908	1.796
Skew	-0.869	-0.892	-0.234
Kurtosis	0.541	0.954	-0.764
Minimum	42	36	0
Maximum	120	121	7
Median	99	97	4
Alpha	0.913	0.921	0.594
SEM	4.077	4.178	1.145
Mean P	0.765	0.756	0.584
Mean Item-Tot.	0.300	0.316	0.540
Mean Biserial	0.480	0.491	0.700
Max Score (Low)	89	87	3
N (Low Group)	122	89	281
Min Score (High)	105	104	5
N (High Group)	126	98	328

The results of the item analyses indicate very similar means, standard deviations, and reliabilities for the two forms of the MC test. The alpha coefficients for Form A and Form B were .913 and .921 respectively indicating excellent reliability. The histograms, as well as the skewness and kurtosis statistics, shown in Figures 5 and 6 also reveal very similar distributions for Form A and Form B. The coefficient alpha for the seven PA exercises was .594 which is acceptable considering the small number of items included.

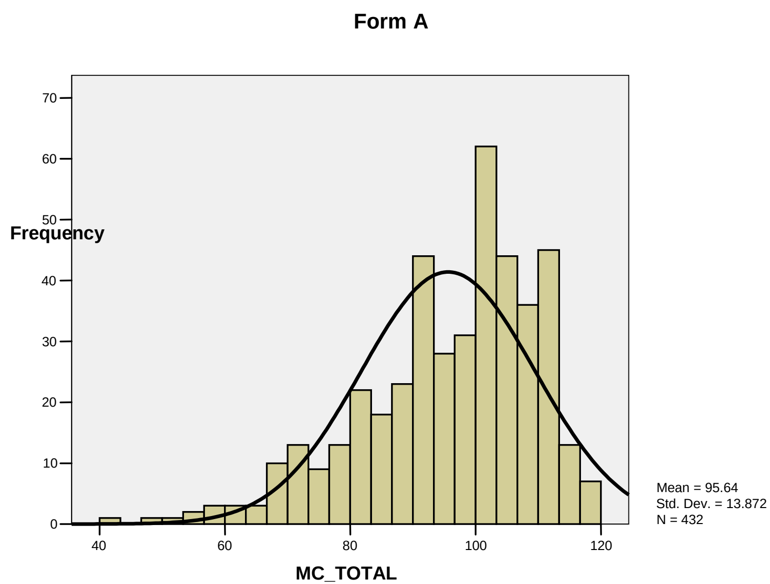

Figure 5. Form A Test Scores Histogram

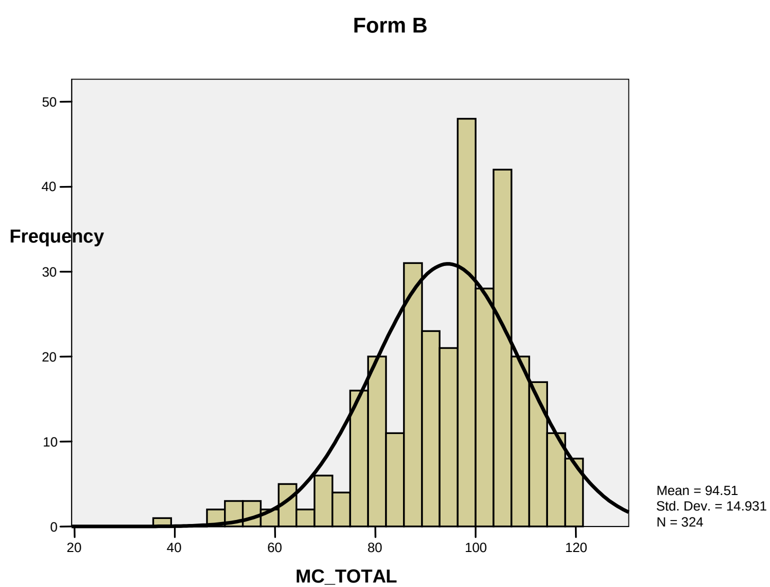

Figure 6. Form B Test Scores Histogram

4.2.2 Item Properties

Item difficulty (proportion correct), and point biserial correlations were calculated for each item and exercise. The results for Form A and Form B are shown in Tables 8 and 9 respectively. For Form A, seventeen of items had point biserial correlations below .20 (items 8, 13, 14, 22, 36, 47, 52, 65, 69, 104, 105, 107, 109, 110, 119, 125, and 126). For Form B, thirteen of items (items 13, 19, 32, 33, 41, 56, 74, 83, 95, 97, 108, 110, and 120) had point biserial correlations below .20. The low point biserial correlations of the non-loading items could be an indication of flawed or ineffective items. The majority of items identified as having low point biserial correlations also had high item difficulty indices (above .80). The high item difficulties suggest that these items may have been too easy for this sample of examinees.

Table 8. Form A Items Statistics

Subsection	Item	Prop. Correct	P.B.
Mechanical	***I3***	0.74	0.28
Mechanical	***I4***	0.62	0.32
Mechanical	***I5***	0.80	0.43
Mechanical	***I6***	0.96	0.21
Mechanical	***I7***	0.47	0.46
Mechanical	***I8***	0.88	0.17
Mechanical	***I9***	0.72	0.41
Mechanical	***I10***	0.68	0.51
Mechanical	***I11***	0.91	0.27
Mechanical	***I12***	0.84	0.47
Mechanical	***I13***	0.63	0.11
Mechanical	***I14***	0.92	0.17
Mechanical	***I15***	0.94	0.27
Mechanical	***I16***	0.94	0.23
Mechanical	***I17***	0.79	0.49
Mechanical	***I18***	0.77	0.46
Mechanical	***I19***	0.87	0.25
Mechanical	***I20***	0.83	0.31
Mechanical	***I21***	0.94	0.28
Mechanical	***I22***	0.37	0.08
Electrical	***I23***	0.80	0.40
Electrical	***I24***	0.97	0.17
Electrical	***I25***	0.84	0.16
Electrical	***I26***	0.97	0.15
Electrical	***I27***	0.75	0.09
Electrical	***I28***	0.33	0.10
Electrical	***I29***	0.66	0.49
Electrical	***I30***	0.85	0.21
Electrical	***I31***	0.23	0.14
Electrical	***I32***	0.89	0.36
Electrical	***I33***	0.70	-0.01
Electrical	***I34***	0.54	0.32
Electrical	***I35***	0.87	0.40
Electrical	***I36***	0.93	0.14
Electrical	***I37***	0.53	0.49
Electrical	***I38***	0.97	0.24
Electrical	***I39***	0.97	0.32
Electrical	***I40***	0.85	0.50
Electrical	***I41***	0.93	0.25
Electrical	***I42***	0.82	0.29
Electrical	***I43***	0.39	0.28
Electrical	***I44***	0.94	0.20
Electrical	***I45***	0.49	0.43
Electrical	***I46***	0.39	0.29
Electrical	***I47***	0.63	-0.11
Electrical	***I48***	0.44	0.36
Electrical	***I49***	0.57	0.40
Electrical	***I50***	0.97	0.20
Electrical	***I51***	0.44	0.29
Electrical	***I52***	0.36	0.24
Hydraulics	***I53***	0.85	0.24
Hydraulics	***I54***	0.85	0.53
Hydraulics	***I55***	0.95	0.31
Hydraulics	***I56***	0.84	0.46
Hydraulics	***I57***	0.83	0.33
Hydraulics	***I58***	0.67	0.28
Hydraulics	***I59***	0.69	0.21
Hydraulics	***I60***	0.70	0.48
Hydraulics	***I61***	0.50	0.29
Hydraulics	***I62***	0.50	0.31
Hand Tools	***I63***	0.91	0.42
Hand Tools	***I64***	0.96	0.25
Hand Tools	***I65***	0.98	0.04
Hand Tools	***I66***	0.68	0.24
Hand Tools	***I67***	0.77	0.32
Hand Tools	***I68***	0.71	0.28
Hand Tools	***I69***	0.98	0.12
Hand Tools	***I70***	0.84	0.44

Table 8 (continued).

Subsection	Item	Prop. Correct	P.B.
Math/Stats	***I71***	0.83	0.48
Math/Stats	***I72***	0.98	0.19
Math/Stats	***I73***	0.91	0.25
Math/Stats	***I74***	0.96	0.30
Math/Stats	***I75***	0.97	0.29
Math/Stats	***I76***	0.83	0.25
Math/Stats	***I77***	0.95	0.27
Math/Stats	***I78***	0.95	0.38
Math/Stats	***I79***	0.78	0.33
Math/Stats	***I80***	0.99	0.13
Problem Solving	***I81***	0.74	0.55
Problem Solving	***I82***	0.76	0.53
Problem Solving	***I83***	0.97	0.29
Problem Solving	***I84***	0.80	0.40
Problem Solving	***I85***	0.85	0.48
Problem Solving	***I86***	0.81	0.45
Problem Solving	***I87***	0.78	0.27
Problem Solving	***I88***	0.66	0.37
Problem Solving	***I89***	0.91	0.37
Problem Solving	***I90***	0.49	0.35
Problem Solving	***I91***	0.94	0.41
Problem Solving	***I92***	0.81	0.57
Problem Solving	***I93***	0.87	0.31
Problem Solving	***I94***	0.94	0.33
Problem Solving	***I95***	0.78	0.38
Problem Solving	***I96***	0.93	0.36
Problem Solving	***I97***	0.86	0.30
Problem Solving	***I98***	0.70	0.43
Problem Solving	***I99***	0.98	0.26
Problem Solving	***I100***	0.96	0.39
Problem Solving	***I101***	0.94	0.33
Problem Solving	***I102***	0.90	0.37
Problem Solving	***I103***	0.91	0.31
Problem Solving	***I104***	0.93	0.18

Subsection	Item	Prop. Correct	P.B.
Problem Solving	***I105***	0.94	0.18
Problem Solving	***I106***	0.60	0.48
Pneumatics	***I107***	0.40	0.02
Pneumatics	***I108***	0.91	0.34
Pneumatics	***I109***	0.34	-0.04
Pneumatics	***I110***	0.63	0.05
Pneumatics	***I111***	0.81	0.33
Pneumatics	***I112***	0.57	0.21
Pneumatics	***I113***	0.63	0.30
Pneumatics	***I114***	0.66	0.26
Pneumatics	***I115***	0.87	0.44
Pneumatics	***I116***	0.83	0.39
Pneumatics	***I117***	0.27	0.28
Pneumatics	***I118***	0.52	0.51
Pneumatics	***I119***	0.35	0.08
Pneumatics	***I120***	0.88	0.37
Pneumatics	***I121***	0.80	0.31
Pneumatics	***I122***	0.69	0.44
Pneumatics	***I123***	0.85	0.27
Pneumatics	***I124***	0.81	0.32
Pneumatics	***I125***	0.56	0.11
Pneumatics	***I126***	0.66	0.14
Pneumatics	***I127***	0.90	0.25
PA 1	***I128***	0.87	0.38
PA 2	***I129***	0.61	0.39
PA 3	***I130***	0.59	0.32
PA 4	***I131***	0.70	0.42
PA 5	***I132***	0.54	0.48
PA 6	***I133***	0.43	0.36
PA 7	***I134***	0.45	0.31

Table 9. Form B Item Statistics

Subsection	Item	Prop. Correct	P.B.
Mechanical	***I3***	0.71	0.32
Mechanical	***I4***	0.71	0.31
Mechanical	***I5***	0.36	0.27
Mechanical	***I6***	0.94	0.33
Mechanical	***I7***	0.45	0.35
Mechanical	***I8***	0.67	0.42
Mechanical	***I9***	0.65	0.31
Mechanical	***I10***	0.70	0.23
Mechanical	***I11***	0.90	0.37
Mechanical	***I12***	0.97	0.22
Mechanical	***I13***	0.51	0.17
Mechanical	***I14***	0.74	0.47
Mechanical	***I15***	0.93	0.25
Mechanical	***I16***	0.83	0.29
Mechanical	***I17***	0.81	0.25
Mechanical	***I18***	0.61	0.34
Mechanical	***I19***	0.28	0.07
Mechanical	***I20***	0.47	0.24
Mechanical	***I21***	0.80	0.31
Mechanical	***I22***	0.90	0.33
Electrical	***I23***	0.96	0.20
Electrical	***I24***	0.66	0.51
Electrical	***I25***	0.61	0.45
Electrical	***I26***	0.80	0.32
Electrical	***I27***	0.90	0.26
Electrical	***I28***	0.53	0.25
Electrical	***I29***	0.81	0.53
Electrical	***I30***	0.96	0.33
Electrical	***I31***	0.90	0.28
Electrical	***I32***	0.79	0.17
Electrical	***I33***	0.78	0.08
Electrical	***I34***	0.79	0.46
Electrical	***I35***	0.90	0.44
Electrical	***I36***	0.71	0.48

Subsection	Item	Prop. Correct	P.B.
Electrical	***I37***	0.57	0.47
Electrical	***I38***	0.91	0.42
Electrical	***I39***	0.90	0.32
Electrical	***I40***	0.68	0.38
Electrical	***I41***	0.54	0.17
Electrical	***I42***	0.56	0.32
Electrical	***I43***	0.88	0.45
Electrical	***I44***	0.80	0.35
Electrical	***I45***	0.49	0.23
Electrical	***I46***	0.90	0.34
Electrical	***I47***	0.76	0.53
Electrical	***I48***	0.73	0.44
Electrical	***I49***	0.37	0.37
Electrical	***I50***	0.90	0.51
Electrical	***I51***	0.84	0.27
Electrical	***I52***	0.46	0.38
Hydraulics	***I53***	0.51	0.27
Hydraulics	***I54***	0.87	0.26
Hydraulics	***I55***	0.57	0.36
Hydraulics	***I56***	0.90	0.11
Hydraulics	***I57***	0.45	0.24
Hydraulics	***I58***	0.64	0.22
Hydraulics	***I59***	0.87	0.35
Hydraulics	***I60***	0.90	0.35
Hydraulics	***I61***	0.96	0.30
Hydraulics	***I62***	0.64	0.32
Hand Tools	***I63***	0.88	0.32
Hand Tools	***I64***	0.26	0.31
Hand Tools	***I65***	0.66	0.08
Hand Tools	***I66***	0.45	0.26
Hand Tools	***I67***	0.69	0.24
Hand Tools	***I68***	0.56	0.31
Hand Tools	***I69***	0.84	0.31
Hand Tools	***I70***	0.91	0.29

Table 9 (continued).

Subsection	Item	Prop. Correct	P.B.
Math/Stats	***I71***	0.97	0.25
Math/Stats	***I72***	0.97	0.35
Math/Stats	***I73***	0.82	0.40
Math/Stats	***I74***	0.82	0.13
Math/Stats	***I75***	0.82	0.45
Math/Stats	***I76***	0.86	0.43
Math/Stats	***I77***	0.90	0.34
Math/Stats	***I78***	0.94	0.42
Math/Stats	***I79***	0.56	0.28
Math/Stats	***I80***	0.96	0.25
Problem Solving	***I81***	0.76	0.34
Problem Solving	***I82***	0.86	0.23
Problem Solving	***I83***	0.37	0.13
Problem Solving	***I84***	0.84	0.40
Problem Solving	***I85***	0.90	0.41
Problem Solving	***I86***	0.77	0.28
Problem Solving	***I87***	0.84	0.22
Problem Solving	***I88***	0.98	0.24
Problem Solving	***I89***	0.89	0.31
Problem Solving	***I90***	0.95	0.33
Problem Solving	***I91***	0.89	0.34
Problem Solving	***I92***	0.79	0.50
Problem Solving	***I93***	0.94	0.30
Problem Solving	***I94***	0.86	0.47
Problem Solving	***I95***	0.80	0.16
Problem Solving	***I96***	0.79	0.33
Problem Solving	***I97***	0.84	0.14
Problem Solving	***I98***	0.96	0.27
Problem Solving	***I99***	0.92	0.30
Problem Solving	***I100***	0.95	0.34
Problem Solving	***I101***	0.83	0.42
Problem Solving	***I102***	0.96	0.31
Problem Solving	***I103***	0.95	0.35
Problem Solving	***I104***	0.88	0.38

Subsection	Item	Prop. Correct	P.B.
Problem Solving	***I105***	0.64	0.44
Problem Solving	***I106***	0.87	0.27
Pneumatics	***I107***	0.83	0.34
Pneumatics	***I108***	0.75	0.16
Pneumatics	***I109***	0.71	0.34
Pneumatics	***I110***	0.68	0.11
Pneumatics	***I111***	0.68	0.25
Pneumatics	***I112***	0.89	0.42
Pneumatics	***I113***	0.46	0.30
Pneumatics	***I114***	0.42	0.41
Pneumatics	***I115***	0.70	0.50
Pneumatics	***I116***	0.60	0.31
Pneumatics	***I117***	0.90	0.48
Pneumatics	***I118***	0.53	0.36
Pneumatics	***I119***	0.68	0.36
Pneumatics	***I120***	0.21	0.05
Pneumatics	***I121***	0.78	0.24
Pneumatics	***I122***	0.77	0.25
Pneumatics	***I123***	0.89	0.42
Pneumatics	***I124***	0.82	0.30
Pneumatics	***I125***	0.92	0.22
Pneumatics	***I126***	0.92	0.38
Pneumatics	***I127***	0.89	0.20
PA 1	***I128***	0.78	0.40
PA 2	***I129***	0.57	0.34
PA 3	***I130***	0.54	0.29
PA 4	***I131***	0.63	0.37
PA 5	***I132***	0.62	0.38
PA 6	***I133***	0.36	0.43
PA 7	***I134***	0.48	0.37

Item statistics for the seven PA exercises are shown in Table 10. The proportion correct statistics reveal that the most difficult of the seven exercises was PA Item 6 (Component Connection) and the easiest was PA Item 1 (Plate Alignment). The results reveal relatively high point biserial correlations for the seven exercises.

Table 10. PA Item Statistics

PA Item	Prop. Correct	P.B.
1	0.83	0.49
2	0.59	0.57
3	0.57	0.46
4	0.67	0.59
5	0.57	0.59
6	0.40	0.59
7	0.46	0.49

4.3 RESEARCH QUESTION 3 - WHAT IS THE FACTOR STRUCTURE OF BOTH FORM A AND FORM B WITH THE PA EXERCISES INCLUDED?

4.3.1 CFA Results

The computer program Mplus Version 3.11 for Windows was used to conduct a confirmatory factor analysis on the MC items combined with the PA exercises. Three different competing models were evaluated. Model 1 hypothesizes that the one-factor model will fit the MC test combined with the PA exercises since both portions of the test were designed to measure a construct labeled as *applied mechanical knowledge*.

Model 2 hypothesizes that there are two different dimensions underlying the test scores. The two dimensions consist of separate dimensions for the MC portion and the PA exercises.

Model 3 hypothesizes that there are eight different dimensions underlying the test scores. The eight dimensions consist of the seven content areas of the MC portion (Mechanical, Electrical, Hydraulics, Hand/Measuring Tools, Math/Statistics, Problem Solving, and Pneumatics) plus one additional dimension for the PA exercises.

4.3.1.1 CFA Fit Statistics Because assumptions for the chi-square test are generally violated when factor analysis is conducted, it was decided to accept the common practice of dismissing the chi-square test as a formal hypothesis test and instead rely on other methods to assess fit of the model to the data.

CFA goodness-of-fit indices implemented by Mplus include the Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI). These two comparative fit indices measure the improvement of fit by comparing the hypothesized model with a more restricted baseline model where the observed variables, with variances to be estimated are mutually uncorrelated (Bentler & Bonett, 1980). Both the CFI and TLI have a 0-1 range, tend toward 1 for a correctly specified model, and have a recommended cutoff value of 0.95 (Hu & Bentler, 1999).

The Root-mean-square Error of Approximation (RMSEA) is a measure of the residual variances and covariances, which quantifies the error of approximation of the population data by the model (Loehlin, 1998). Small values of the RMSEA indicate fit, while an RMSEA value of zero would indicate perfect fit. RMSEA values less than .05 indicate very good fit of the factor model to the data, values between 0.05 and 0.08 indicate moderate fit, and those between 0.08 and 0.1 indicate relatively poor fit (Browne & Cudeck, 1993). According to Hu and Bentler (1999), RMSEA values below .06 indicate satisfactory fit of the model to the data.

The Standardized Root-mean-square Residual (SRMR) and the Weighted Root-mean-square Residual (WRMR) measure the average differences between the sample and estimated population variances and covariances (Yu, 2002). The SRMR has a 0-1 range with a recommended cutoff value close to 0.08 (Hu & Bentler, 1999). Yu (2002) found 1.0 to be an acceptable cutoff for the WRMR for both continuous and dichotomous outcomes.

The CFA models were evaluated for both Form A and Form B. The CFA fit statistics for the three competing models are shown in Table 11. It was hypothesized the CFA results for both forms will be the same due to the fact that Form A and Form B were designed to be alternate equivalent versions of the same test and, as such, both forms share the exact same test blueprint.

Table 11. CFA Fit Statistics (WLSMV)

	Chi-square	p-value	CFI	TLI	RMSEA	SRMR	WRMR
Model 1 - One Factor for Form A MC & PA	314.527	0.0000	0.885	0.907	0.035	0.119	1.112
Model 2 - Two Factors for Form A MC & PA	313.749	0.0000	0.886	0.908	0.035	0.119	1.110
Model 3 - Eight Factors for Form A & PA	*	*	*	*	*	*	*
Model 1 - One Factor for Form B MC & PA	434.889	0.0000	0.686	0.757	0.070	0.151	1.449
Model 2 - Two Factors for Form B MC & PA	433.928	0.0000	0.687	0.758	0.070	0.151	1.448
Model 3 - Eight Factors for Form B & PA	322.497	0.0000	0.816	0.856	0.054	0.139	1.255

*NO CONVERGENCE. NUMBER OF ITERATIONS EXCEEDED.

4.3.1.2 Form A CFA With the exception of the RMSEA fit statistics for Model 1 and Model 2, the results failed to demonstrate fit of the model to the data. The RMSEA value was identical 0.035 for both Model 1 and Model 2. There was no convergence for Model 3 as the number of iterations was exceeded.

4.3.1.3 Form B CFA With the exception of the RMSEA fit statistic for Model 3 (0.054), the results for Model 1, Model 2, and Model 3 failed to demonstrate fit of the model to the data.

4.3.2 EFA Results

Further analysis of the factorial structure of the two forms of the L2/L3 Production Technician Test combined with the 7 PA exercises was undertaken by conducting several exploratory factor analyses using Mplus version 3.11. The factor extraction method, weighted least squares means and variances adjusted (WLSMV) was used since the L2/L3 Production Technician test data are categorical. The oblique rotation method, Promax, was initially used since this method allows the factors to correlate and it was expected that there would be some correlation among factors. Although initially eight factors were extracted for both Form A and Form B, the scree plots and fit statistics for the two forms revealed that an examination of the one and two factor solutions was most appropriate.

4.3.2.1 Form A EFA The number of factors to retain for rotation was determined by an analysis of the graph of the eigenvalues or scree plot. The eigenvalues explained by each factor are plotted in Figure 7. An examination of the scree plot for Form A combined with the 7 PA exercises suggested that a one factor structure was acceptable.

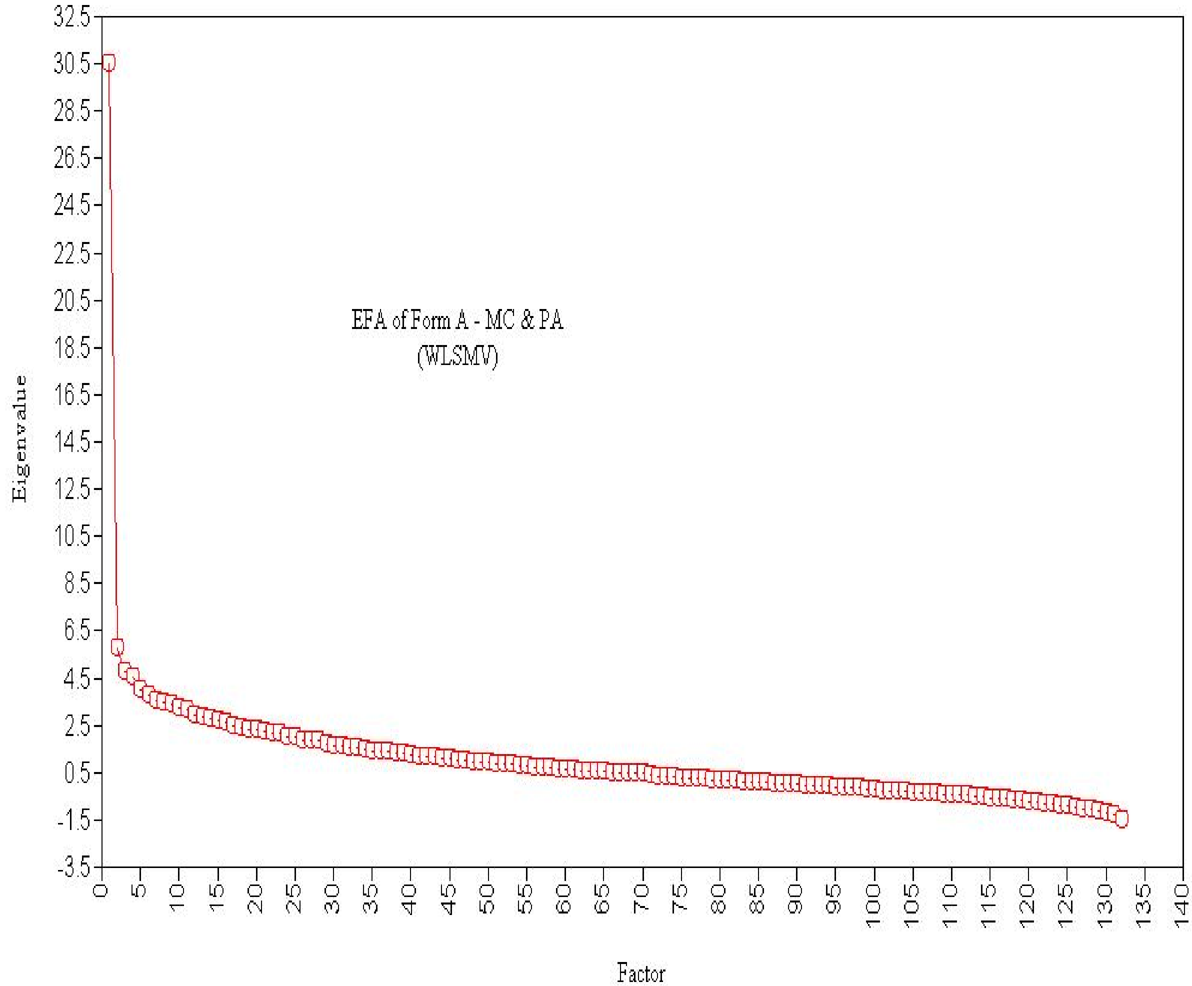

Figure 7. Scree Plot for Form A

4.3.2.2 Form A EFA Fit Statistics Because assumptions for the chi-square test are generally violated when factor analysis is conducted, it was decided to accept the common practice of dismissing the chi-square test as a formal hypothesis test and instead rely on other methods to assess fit of the model to the data.

EFA goodness of fit indices which are implemented by Mplus include the root mean square residual (RMR) and the root mean square error of approximation (RMSEA). The root mean square residual (RMR) is the square root of the average square residuals. It summarizes the differences between the observed and expected covariances given the model. Larger values indicate less fit between the model and the data. According to Hu and Bentler (1999) RMR should be below .08 with lower values indicating better fit of the model.

The EFA fit statistics for Form A combined with the PA exercises are shown in Table 12.

Table 12. Form A with PA EFA Statistics Using WLSMV

Number of Factors Extracted	Chi-square	p-value	RMSEA	RMR
1	314.527	0.0000	0.035	0.1195
2	285.687	0.0002	0.030	0.1141

While the Chi-square and RMR fit statistics did not indicate acceptable fit of either the one or two factor solutions, the RMSEA fit statistics were well below the .06 cutoff criteria for both models indicating acceptable fit.

4.3.2.3 Form A One-Factor Solution The factor loadings for the first factor along with the corresponding item statistics (proportion correct or p_i, and point biserial correlations) and subsections are presented in Table 13. Regarding factor loadings, Gorsuch (1983) reports that an absolute value of .3 is commonly used as the minimum loading for interpretation. All but 28 items (items 8, 13, 14, 22, 25, 26, 27, 28, 30, 31, 33, 36, 47, 52, 59, 65, 66, 69, 104, 105, 107, 109, 110, 112, 119, 125, 126, and 127) loaded above .3 on factor 1. Of the 28 items that did not load above .3 on factor one, all had point biserial correlations less than or equal to .25. Twenty-two of the non-loading items had point biserial correlations below .20 (items 8, 13, 14, 22, 25, 26, 27, 28, 31, 33, 36, 47, 65, 69, 104, 105, 107, 109, 110, 119, 125, and 126). The low point biserial correlations of the non-loading items indicate items that add little or no information to the test for employee selection purposes. Many of the items identified as having low point biserial correlations also had high item difficulty indices (above .80). The high item difficulties suggest that these items may have been too easy for this sample of examinees.

Table 13. Form A Factor Loadings for One Factor with Item Statistics

Subsection	Item	Factor One	p_i	P.B.	Subsection	Item	Factor One	p_i	P.B.
Mechanical	***I3***	**-0.345**	0.74	0.28	Electrical	***I37***	**-0.622**	0.53	0.49
Mechanical	***I4***	**-0.393**	0.62	0.32	Electrical	***I38***	**-0.487**	0.97	0.24
Mechanical	***I5***	**-0.588**	0.80	0.43	Electrical	***I39***	**-0.779**	0.97	0.32
Mechanical	***I6***	**-0.431**	0.96	0.21	Electrical	***I40***	**-0.703**	0.85	0.50
Mechanical	***I7***	**-0.568**	0.47	0.46	Electrical	***I41***	**-0.447**	0.93	0.25
Mechanical	***I8***	-0.263	0.88	0.17	Electrical	***I42***	**-0.377**	0.82	0.29
Mechanical	***I9***	**-0.512**	0.72	0.41	Electrical	***I43***	**-0.339**	0.39	0.28
Mechanical	***I10***	**-0.654**	0.68	0.51	Electrical	***I44***	**-0.392**	0.94	0.20
Mechanical	***I11***	**-0.428**	0.91	0.27	Electrical	***I45***	**-0.543**	0.49	0.43
Mechanical	***I12***	**-0.698**	0.84	0.47	Electrical	***I46***	**-0.359**	0.39	0.29
Mechanical	***I13***	-0.084	0.63	0.11	Electrical	***I47***	0.212	0.63	-0.11
Mechanical	***I14***	-0.259	0.92	0.17	Electrical	***I48***	**-0.440**	0.44	0.36
Mechanical	***I15***	**-0.481**	0.94	0.27	Electrical	***I49***	**-0.493**	0.57	0.40
Mechanical	***I16***	**-0.431**	0.94	0.23	Electrical	***I50***	**-0.405**	0.97	0.20
Mechanical	***I17***	**-0.679**	0.79	0.49	Electrical	***I51***	**-0.357**	0.44	0.29
Mechanical	***I18***	**-0.618**	0.77	0.46	Electrical	***I52***	-0.249	0.36	0.24
Mechanical	***I19***	**-0.377**	0.87	0.25	Hydraulics	***I53***	**-0.319**	0.85	0.24
Mechanical	***I20***	**-0.442**	0.83	0.31	Hydraulics	***I54***	**-0.755**	0.85	0.53
Mechanical	***I21***	**-0.467**	0.94	0.28	Hydraulics	***I55***	**-0.609**	0.95	0.31
Mechanical	***I22***	-0.065	0.37	0.08	Hydraulics	***I56***	**-0.642**	0.84	0.46
Electrical	***I23***	**-0.548**	0.80	0.40	Hydraulics	***I57***	**-0.455**	0.83	0.33
Electrical	***I24***	**-0.358**	0.97	0.17	Hydraulics	***I58***	**-0.339**	0.67	0.28
Electrical	***I25***	-0.216	0.84	0.16	Hydraulics	***I59***	-0.229	0.69	0.21
Electrical	***I26***	-0.247	0.97	0.15	Hydraulics	***I60***	**-0.618**	0.70	0.48
Electrical	***I27***	-0.087	0.75	0.09	Hydraulics	***I61***	**-0.378**	0.50	0.29
Electrical	***I28***	-0.086	0.33	0.10	Hydraulics	***I62***	**-0.374**	0.50	0.31
Electrical	***I29***	**-0.611**	0.66	0.49	Hand Tools	***I63***	**-0.629**	0.91	0.42
Electrical	***I30***	-0.288	0.85	0.21	Hand Tools	***I64***	**-0.480**	0.96	0.25
Electrical	***I31***	-0.156	0.23	0.14	Hand Tools	***I65***	-0.035	0.98	0.04
Electrical	***I32***	**-0.563**	0.89	0.36	Hand Tools	***I66***	-0.248	0.68	0.24
Electrical	***I33***	0.050	0.70	-0.01	Hand Tools	***I67***	**-0.459**	0.77	0.32
Electrical	***I34***	**-0.380**	0.54	0.32	Hand Tools	***I68***	**-0.334**	0.71	0.28
Electrical	***I35***	**-0.600**	0.87	0.40	Hand Tools	***I69***	-0.227	0.98	0.12
Electrical	***I36***	-0.114	0.93	0.14	Hand Tools	***I70***	**-0.620**	0.84	0.44

Table 13 (continued).

Subsection	Item	Factor One	p_i	P.B.
Math/Stats	***I71***	**-0.701**	0.83	0.48
Math/Stats	***I72***	**-0.535**	0.98	0.19
Math/Stats	***I73***	**-0.429**	0.91	0.25
Math/Stats	***I74***	**-0.607**	0.96	0.30
Math/Stats	***I75***	**-0.688**	0.97	0.29
Math/Stats	***I76***	**-0.325**	0.83	0.25
Math/Stats	***I77***	**-0.533**	0.95	0.27
Math/Stats	***I78***	**-0.754**	0.95	0.38
Math/Stats	***I79***	**-0.451**	0.78	0.33
Math/Stats	***I80***	**-0.444**	0.99	0.13
Prob. Solve	***I81***	**-0.693**	0.74	0.55
Prob. Solve	***I82***	**-0.707**	0.76	0.53
Prob. Solve	***I83***	**-0.643**	0.97	0.29
Prob. Solve	***I84***	**-0.588**	0.80	0.40
Prob. Solve	***I85***	**-0.705**	0.85	0.48
Prob. Solve	***I86***	**-0.643**	0.81	0.45
Prob. Solve	***I87***	**-0.367**	0.78	0.27
Prob. Solve	***I88***	**-0.476**	0.66	0.37
Prob. Solve	***I89***	**-0.623**	0.91	0.37
Prob. Solve	***I90***	**-0.461**	0.49	0.35
Prob. Solve	***I91***	**-0.749**	0.94	0.41
Prob. Solve	***I92***	**-0.790**	0.81	0.57
Prob. Solve	***I93***	**-0.473**	0.87	0.31
Prob. Solve	***I94***	**-0.618**	0.94	0.33
Prob. Solve	***I95***	**-0.522**	0.78	0.38
Prob. Solve	***I96***	**-0.620**	0.93	0.36
Prob. Solve	***I97***	**-0.419**	0.86	0.30
Prob. Solve	***I98***	**-0.551**	0.70	0.43
Prob. Solve	***I99***	**-0.672**	0.98	0.26
Prob. Solve	***I100***	**-0.786**	0.96	0.39
Prob. Solve	***I101***	**-0.636**	0.94	0.33
Prob. Solve	***I102***	**-0.592**	0.90	0.37
Prob. Solve	***I103***	**-0.524**	0.91	0.31
Prob. Solve	***I104***	-0.284	0.93	0.18

Subsection	Item	Factor One	p_i	P.B.
Prob. Solve	***I105***	-0.273	0.94	0.18
Prob. Solve	***I106***	**-0.611**	0.60	0.48
Pneumatics	***I107***	0.009	0.40	0.02
Pneumatics	***I108***	**-0.521**	0.91	0.34
Pneumatics	***I109***	0.121	0.34	-0.04
Pneumatics	***I110***	-0.035	0.63	0.05
Pneumatics	***I111***	**-0.429**	0.81	0.33
Pneumatics	***I112***	-0.222	0.57	0.21
Pneumatics	***I113***	**-0.368**	0.63	0.30
Pneumatics	***I114***	**-0.306**	0.66	0.26
Pneumatics	***I115***	**-0.647**	0.87	0.44
Pneumatics	***I116***	**-0.550**	0.83	0.39
Pneumatics	***I117***	**-0.375**	0.27	0.28
Pneumatics	***I118***	**-0.688**	0.52	0.51
Pneumatics	***I119***	-0.034	0.35	0.08
Pneumatics	***I120***	**-0.525**	0.88	0.37
Pneumatics	***I121***	**-0.359**	0.80	0.31
Pneumatics	***I122***	**-0.541**	0.69	0.44
Pneumatics	***I123***	**-0.382**	0.85	0.27
Pneumatics	***I124***	**-0.414**	0.81	0.32
Pneumatics	***I125***	-0.125	0.56	0.11
Pneumatics	***I126***	-0.111	0.66	0.14
Pneumatics	***I127***	-0.283	0.90	0.25
PA 1	***I128***	**-0.552**	0.87	0.38
PA 2	***I129***	**-0.476**	0.61	0.39
PA 3	***I130***	**-0.381**	0.59	0.32
PA 4	***I131***	**-0.514**	0.70	0.42
PA 5	***I132***	**-0.586**	0.54	0.48
PA 6	***I133***	**-0.459**	0.43	0.36
PA 7	***I134***	**-0.384**	0.45	0.31

4.3.2.4 Form A Two-Factor Solution Mplus version 3.11 was also used to conduct an EFA on Form A combined with the 7 PA exercises with two factors extracted. Although an absolute value of .3 is commonly used as the minimum loading for interpretation, this rule is in fact only appropriate for orthogonal rotations. For an oblique rotation such as Promax, the structure matrix must be computed by multiplying the factor loading matrix by the factor correlation matrix. The factor correlation matrix is shown in Table 14. The Promax factor loadings and the structure coefficients for the first two factors, as well as item statistics and subsections are shown in Table 15. The minimum loading for interpretation is an absolute value of .3 on both the factor loading and structure coefficient.

Table 14. Form A Factor Correlations

1.000	**0.516**
0.516	**1.000**

The two-factor solution was significantly more complex than the one-factor solution, and was not easily interpreted. There was a significant number of crossloading to the extent that the majority of items loaded above .3 on both factors and the corresponding structure coefficients. The results indicated that the two-factor solution was clearly overfactoring. Therefore, the results of the EFA on Form A combined with the PA exercises supported a one-factor solution.

Table 15. Form A Promax Factor Loadings for Two Factors with Structure Coefficients and Item Statistics

		Promax Loadings		**Structure Coefficients**			
Subsection	**Item**	**Factor One**	**Factor Two**	**Factor One**	**Factor Two**	***p*** $_i$	**P.B.**
Mechanical	***I3***	0.250	0.159	**0.332**	0.288	0.74	0.28
Mechanical	***I4***	**0.388**	0.091	**0.435**	0.291	0.62	0.32
Mechanical	***I5***	0.106	**0.541**	**0.385**	**0.596**	0.80	0.43
Mechanical	***I6***	0.158	**0.330**	**0.328**	**0.412**	0.96	0.21
Mechanical	***I7***	**0.542**	0.148	**0.618**	**0.428**	0.47	0.46
Mechanical	***I8***	0.026	0.260	0.160	0.273	0.88	0.17
Mechanical	***I9***	**0.527**	0.101	**0.579**	**0.373**	0.72	0.41
Mechanical	***I10***	**0.500**	0.278	**0.643**	**0.536**	0.68	0.51
Mechanical	***I11***	0.289	0.218	**0.401**	**0.367**	0.91	0.27
Mechanical	***I12***	0.045	**0.702**	**0.407**	**0.725**	0.84	0.47
Mechanical	***I13***	0.173	-0.057	0.144	0.032	0.63	0.11
Mechanical	***I14***	0.063	0.224	0.179	0.257	0.92	0.17
Mechanical	***I15***	**0.302**	0.264	**0.438**	**0.420**	0.94	0.27
Mechanical	***I16***	**-0.356**	**0.758**	0.035	**0.574**	0.94	0.23
Mechanical	***I17***	0.020	**0.702**	**0.382**	**0.712**	0.79	0.49
Mechanical	***I18***	0.118	**0.562**	**0.408**	**0.623**	0.77	0.46
Mechanical	***I19***	0.082	**0.335**	0.255	**0.377**	0.87	0.25
Mechanical	***I20***	0.258	0.258	**0.391**	**0.391**	0.83	0.31
Mechanical	***I21***	-0.122	**0.600**	0.188	**0.537**	0.94	0.28
Mechanical	***I22***	0.034	0.041	0.055	0.059	0.37	0.08
Electrical	***I23***	0.048	**0.546**	**0.330**	**0.571**	0.80	0.40
Electrical	***I24***	0.256	0.177	**0.347**	**0.309**	0.97	0.17
Electrical	***I25***	0.151	0.105	0.205	0.183	0.84	0.16
Electrical	***I26***	0.178	0.121	0.240	0.213	0.97	0.15
Electrical	***I27***	-0.097	0.176	-0.006	0.126	0.75	0.09
Electrical	***I28***	0.207	-0.085	0.163	0.022	0.33	0.10
Electrical	***I29***	**0.575**	0.167	**0.661**	**0.464**	0.66	0.49
Electrical	***I30***	0.287	0.066	**0.321**	0.214	0.85	0.21
Electrical	***I31***	0.133	0.054	0.161	0.123	0.23	0.14
Electrical	***I32***	0.288	**0.364**	**0.476**	**0.513**	0.89	0.36
Electrical	***I33***	-0.224	0.138	-0.153	0.022	0.70	-0.01
Electrical	***I34***	**0.430**	0.041	**0.451**	0.263	0.54	0.32
Electrical	***I35***	**0.153**	0.514	**0.418**	0.593	0.87	0.40
Electrical	***I36***	**0.506**	-0.272	**0.366**	-0.011	0.93	0.14
Electrical	***I37***	**0.462**	0.276	**0.604**	0.514	0.53	0.49
Electrical	***I38***	**-0.103**	0.611	**0.212**	0.558	0.97	0.24
Electrical	***I39***	**0.351**	0.542	**0.631**	0.723	0.97	0.32
Electrical	***I40***	**0.368**	0.446	**0.598**	0.636	0.85	0.50

Table 15 (continued).

Subsection	Item	Promax Loadings Factor One	Promax Loadings Factor Two	Structure Coefficients Factor One	Structure Coefficients Factor Two	p_i	P.B.
Electrical	***I41***	0.242	0.277	**0.385**	**0.402**	0.93	0.25
Electrical	***I42***	**0.462**	0.013	**0.469**	0.251	0.82	0.29
Electrical	***I43***	**0.641**	-0.191	**0.542**	0.140	0.39	0.28
Electrical	***I44***	-0.208	**0.585**	0.094	**0.478**	0.94	0.20
Electrical	***I45***	**0.695**	-0.017	**0.686**	**0.342**	0.49	0.43
Electrical	***I46***	0.240	0.182	**0.334**	**0.306**	0.39	0.29
Electrical	***I47***	-0.073	-0.167	-0.159	-0.205	0.63	-0.11
Electrical	***I48***	**0.391**	0.141	**0.464**	**0.343**	0.44	0.36
Electrical	***I49***	**0.540**	0.068	**0.575**	**0.347**	0.57	0.40
Electrical	***I50***	-0.006	**0.434**	0.218	**0.431**	0.97	0.20
Electrical	***I51***	0.123	0.279	0.267	**0.342**	0.44	0.29
Electrical	***I52***	**0.522**	-0.181	**0.429**	0.088	0.36	0.24
Hydraulics	***I53***	**0.450**	-0.039	**0.430**	0.193	0.85	0.24
Hydraulics	***I54***	0.247	**0.602**	**0.558**	**0.729**	0.85	0.53
Hydraulics	***I55***	0.161	**0.517**	**0.428**	**0.600**	0.95	0.31
Hydraulics	***I56***	**0.334**	**0.409**	**0.545**	**0.581**	0.84	0.46
Hydraulics	***I57***	0.140	**0.371**	**0.331**	**0.443**	0.83	0.33
Hydraulics	***I58***	0.279	0.127	**0.345**	0.271	0.67	0.28
Hydraulics	***I59***	0.143	0.126	0.208	0.200	0.69	0.21
Hydraulics	***I60***	**0.529**	0.214	**0.639**	**0.487**	0.70	0.48
Hydraulics	***I61***	0.199	0.239	**0.322**	**0.342**	0.50	0.29
Hydraulics	***I62***	**0.337**	0.116	**0.397**	0.290	0.50	0.31
Hand Tools	***I63***	**0.331**	**0.397**	**0.536**	**0.568**	0.91	0.42
Hand Tools	***I64***	0.237	**0.318**	**0.401**	**0.440**	0.96	0.25
Hand Tools	***I65***	-0.181	0.165	-0.096	0.072	0.98	0.04
Hand Tools	***I66***	**0.430**	-0.101	**0.378**	0.121	0.68	0.24
Hand Tools	***I67***	**0.082**	0.423	**0.300**	0.465	0.77	0.32
Hand Tools	***I68***	**0.233**	0.161	**0.316**	0.281	0.71	0.28
Hand Tools	***I69***	**0.707**	-0.309	**0.548**	0.056	0.98	0.12
Hand Tools	***I70***	**0.426**	0.307	**0.584**	0.527	0.84	0.44
Math/Stats	***I71***	**-0.107**	0.820	**0.316**	0.765	0.83	0.48
Math/Stats	***I72***	**-0.082**	0.652	**0.254**	0.610	0.98	0.19
Math/Stats	***I73***	**-0.057**	0.501	**0.202**	0.472	0.91	0.25
Math/Stats	***I74***	**-0.107**	0.724	**0.267**	0.669	0.96	0.30
Math/Stats	***I75***	**-0.162**	0.840	**0.271**	0.756	0.97	0.29
Math/Stats	***I76***	**0.056**	0.301	**0.211**	0.330	0.83	0.25
Math/Stats	***I77***	**0.132**	0.457	**0.368**	0.525	0.95	0.27
Math/Stats	***I78***	**0.089**	0.719	**0.460**	0.765	0.95	0.38
Math/Stats	***I79***	**0.094**	0.404	**0.302**	0.453	0.78	0.33
Math/Stats	***I80***	**-0.044**	0.504	**0.216**	0.481	0.99	0.13
Problem Solving	***I81***	**0.227**	0.552	**0.512**	0.669	0.74	0.55
Problem Solving	***I82***	**0.358**	0.459	**0.595**	0.644	0.76	0.53
Problem Solving	***I83***	**0.093**	0.597	**0.401**	0.645	0.97	0.29

Table 15 (continued).

Subsection	Item	Promax Loadings Factor One	Promax Loadings Factor Two	Structure Coefficients Factor One	Structure Coefficients Factor Two	p_i	P.B.
Problem Solving	***I84***	-0.063	**0.676**	0.286	**0.643**	0.80	0.40
Problem Solving	***I85***	0.178	**0.605**	**0.490**	**0.697**	0.85	0.48
Problem Solving	***I86***	0.064	**0.628**	**0.388**	**0.661**	0.81	0.45
Problem Solving	***I87***	0.016	**0.380**	0.212	**0.388**	0.78	0.27
Problem Solving	***I88***	0.043	**0.475**	0.288	**0.497**	0.66	0.37
Problem Solving	***I89***	-0.023	**0.675**	**0.325**	**0.663**	0.91	0.37
Problem Solving	***I90***	0.163	**0.358**	**0.348**	**0.442**	0.49	0.35
Problem Solving	***I91***	-0.080	**0.841**	**0.354**	**0.800**	0.94	0.41
Problem Solving	***I92***	0.047	**0.793**	**0.456**	**0.817**	0.81	0.57
Problem Solving	***I93***	-0.157	**0.628**	0.167	**0.547**	0.87	0.31
Problem Solving	***I94***	-0.109	**0.727**	0.266	**0.671**	0.94	0.33
Problem Solving	***I95***	-0.020	**0.573**	0.276	**0.563**	0.78	0.38
Problem Solving	***I96***	0.041	**0.625**	**0.364**	**0.646**	0.93	0.36
Problem Solving	***I97***	0.250	0.240	**0.374**	**0.369**	0.86	0.30
Problem Solving	***I98***	0.089	**0.516**	**0.355**	**0.562**	0.70	0.43
Problem Solving	***I99***	0.067	**0.648**	**0.401**	**0.683**	0.98	0.26
Problem Solving	***I100***	0.073	**0.775**	**0.473**	**0.813**	0.96	0.39
Problem Solving	***I101***	-0.007	**0.679**	**0.343**	**0.675**	0.94	0.33
Problem Solving	***I102***	-0.179	**0.769**	0.218	**0.677**	0.90	0.37
Problem Solving	***I103***	0.014	**0.547**	0.296	**0.554**	0.91	0.31
Problem Solving	***I104***	0.229	0.115	0.288	0.233	0.93	0.18
Problem Solving	***I105***	-0.295	**0.532**	-0.020	**0.380**	0.94	0.18
Problem Solving	***I106***	0.289	**0.413**	**0.502**	**0.562**	0.60	0.48
Pneumatics	***I107***	-0.085	0.063	-0.052	0.019	0.40	0.02
Pneumatics	***I108***	0.230	**0.367**	**0.419**	**0.486**	0.91	0.34
Pneumatics	***I109***	0.198	**-0.300**	0.043	-0.198	0.34	-0.04
Pneumatics	***I110***	0.032	0.011	0.038	0.028	0.63	0.05
Pneumatics	***I111***	**0.357**	0.160	**0.440**	**0.344**	0.81	0.33
Pneumatics	***I112***	**0.309**	-0.026	0.296	0.133	0.57	0.21
Pneumatics	***I113***	0.116	0.298	0.270	**0.358**	0.63	0.30
Pneumatics	***I114***	0.257	0.111	**0.314**	0.244	0.66	0.26
Pneumatics	***I115***	-0.047	**0.723**	**0.326**	**0.699**	0.87	0.44
Pneumatics	***I116***	0.102	**0.504**	**0.362**	**0.557**	0.83	0.39
Pneumatics	***I117***	**0.324**	0.129	**0.391**	0.296	0.27	0.28
Pneumatics	***I118***	0.146	**0.613**	**0.462**	**0.688**	0.52	0.51
Pneumatics	***I119***	0.197	-0.132	0.129	-0.030	0.35	0.08
Pneumatics	***I120***	0.172	**0.418**	**0.388**	**0.507**	0.88	0.37
Pneumatics	***I121***	**0.743**	-0.247	**0.616**	0.136	0.80	0.31
Pneumatics	***I122***	**0.603**	0.067	**0.638**	**0.378**	0.69	0.44
Pneumatics	***I123***	0.111	**0.316**	0.274	**0.373**	0.85	0.27
Pneumatics	***I124***	0.265	0.222	**0.380**	**0.359**	0.81	0.32
Pneumatics	***I125***	0.077	0.070	0.113	0.110	0.56	0.11
Pneumatics	***I126***	**0.311**	-0.145	0.236	0.015	0.66	0.14

Table 15 (continue)

Subsection	Item	Promax Loadings Factor One	Promax Loadings Factor Two	Structure Coefficients Factor One	Structure Coefficients Factor Two	p_i	P.B.
Pneumatics	***I127***	**0.695**	-0.246	**0.568**	0.113	0.90	0.25
PA 1	***I128***	0.212	**0.414**	**0.426**	**0.523**	0.87	0.38
PA 2	***I129***	0.239	**0.310**	**0.399**	**0.433**	0.61	0.39
PA 3	***I130***	0.131	0.298	0.285	**0.366**	0.59	0.32
PA 4	***I131***	**0.363**	0.246	**0.490**	**0.433**	0.70	0.42
PA 5	***I132***	**0.449**	0.250	**0.578**	**0.482**	0.54	0.48
PA 6	***I133***	0.197	**0.327**	**0.366**	**0.429**	0.43	0.36
PA 7	***I134***	0.213	0.232	**0.333**	**0.342**	0.45	0.31

4.3.2.5 Form B EFA In order to determine the factorial structure of the L2/L3 Production Technician Test Form B combined with the 7 PA exercises, an EFA was conducted using Mplus version 3.11.determine the factor structure of the test. The eigenvalues explained by each factor are plotted in Figure 8. The scree plot for Form B with the 7 PA exercises was similar to the scree plot for Form A, suggesting that a one-factor structure was most appropriate. However, because the second eigenvalue in the Form B scree plot was considerably larger than the second eigenvalue in the Form A scree plot (9.151 versus 5.758 respectively), a two-factor solution was initially given more consideration for Form B.

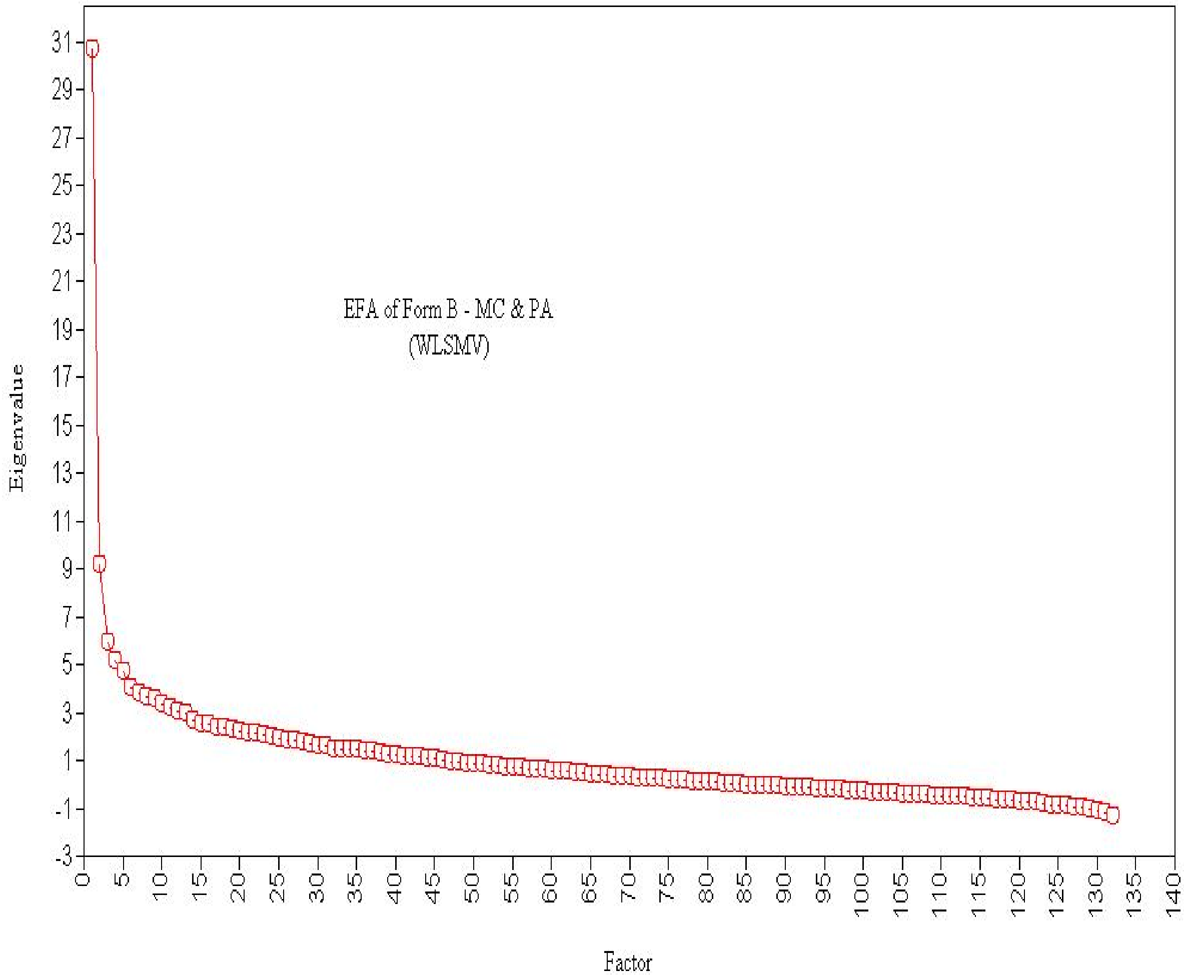

Figure 8. Scree Plot for Form B

4.3.2.6 Form B EFA Fit Statistics The EFA fit statistics for Form B combined with the PA exercises are shown Table 16. While the Chi-square and RMR fit statistics did not indicate acceptable fit of either the one- or two-factor solutions, the RMSEA fit statistic for the two-factor solution indicated acceptable fit. The RMSEA value of .070 indicated moderate fit of the model for the one-factor solution.

Table 16. Form B with PA EFA Fit Statistics Using WLSMV

Number of Factors Extracted	Chi-square	p-value	RMSEA	RMR
1	434.889	0.0000	0.070	0.1523
2	234.904	0.0002	0.037	0.1202

4.3.2.7 Form B One-Factor Solution The factor loadings for the first factor are shown in Table 17. For the one-factor solution, 21 items (items 10, 13, 19, 20, 28, 32, 33, 41, 45, 56, 67, 74, 83, 95, 97, 108, 110, 120, 121, 122, and 127) did not load above .3 on the first factor. All of the 21 non-loading items had point biserial correlations less than or equal to .25. Thirteen of the non-loading items (items 13, 19, 32, 33, 41, 56, 74, 83, 95, 97, 108, 110, and 120) had point biserial correlations below .20. The low point biserial correlations of the non-loading items indicate items that add little or no information to the test for employee selection purposes.

Table 17. Form B Factor Loadings for One Factor with Item Statistics

Subsection	Item	Factor One	p_i	P.B.	Subsection	Item	Factor One	p_i	P.B.
Mechanical	*I3*	**-0.383**	0.71	0.32	Electrical	*I37*	**-0.594**	0.57	0.47
Mechanical	*I4*	**-0.353**	0.71	0.31	Electrical	*I38*	**-0.681**	0.91	0.42
Mechanical	*I5*	**-0.305**	0.36	0.27	Electrical	*I39*	**-0.489**	0.90	0.32
Mechanical	*I6*	**-0.587**	0.94	0.33	Electrical	*I40*	**-0.463**	0.68	0.38
Mechanical	*I7*	**-0.413**	0.45	0.35	Electrical	*I41*	-0.128	0.54	0.17
Mechanical	*I8*	**-0.492**	0.67	0.42	Electrical	*I42*	**-0.340**	0.56	0.32
Mechanical	*I9*	**-0.367**	0.65	0.31	Electrical	*I43*	**-0.650**	0.88	0.45
Mechanical	*I10*	-0.240	0.70	0.23	Electrical	*I44*	**-0.466**	0.80	0.35
Mechanical	*I11*	**-0.581**	0.90	0.37	Electrical	*I45*	-0.264	0.49	0.23
Mechanical	*I12*	**-0.481**	0.97	0.22	Electrical	*I46*	**-0.519**	0.90	0.34
Mechanical	*I13*	-0.167	0.51	0.17	Electrical	*I47*	**-0.684**	0.76	0.53
Mechanical	*I14*	**-0.576**	0.74	0.47	Electrical	*I48*	**-0.550**	0.73	0.44
Mechanical	*I15*	**-0.453**	0.93	0.25	Electrical	*I49*	**-0.568**	0.37	0.37
Mechanical	*I16*	**-0.403**	0.83	0.29	Electrical	*I50*	**-0.747**	0.90	0.51
Mechanical	*I17*	**-0.329**	0.81	0.25	Electrical	*I51*	**-0.323**	0.84	0.27
Mechanical	*I18*	**-0.385**	0.61	0.34	Electrical	*I52*	**-0.619**	0.46	0.38
Mechanical	*I19*	-0.061	0.28	0.07	Hydraulics	*I53*	**-0.445**	0.51	0.27
Mechanical	*I20*	-0.254	0.47	0.24	Hydraulics	*I54*	**-0.531**	0.87	0.26
Mechanical	*I21*	**-0.405**	0.80	0.31	Hydraulics	*I55*	**-0.517**	0.57	0.36
Mechanical	*I22*	**-0.480**	0.90	0.33	Hydraulics	*I56*	-0.163	0.90	0.11
Electrical	*I23*	**-0.426**	0.96	0.20	Hydraulics	*I57*	**-0.372**	0.45	0.24
Electrical	*I24*	**-0.641**	0.66	0.51	Hydraulics	*I58*	**-0.881**	0.64	0.22
Electrical	*I25*	**-0.520**	0.61	0.45	Hydraulics	*I59*	**-0.516**	0.87	0.35
Electrical	*I26*	**-0.432**	0.80	0.32	Hydraulics	*I60*	**-0.515**	0.90	0.35
Electrical	*I27*	**-0.374**	0.90	0.26	Hydraulics	*I61*	**-0.597**	0.96	0.30
Electrical	*I28*	-0.248	0.53	0.25	Hydraulics	*I62*	**-0.948**	0.64	0.32
Electrical	*I29*	**-0.723**	0.81	0.53	Hand Tools	*I63*	**-0.521**	0.88	0.32
Electrical	*I30*	**-0.668**	0.96	0.33	Hand Tools	*I64*	**-0.449**	0.26	0.31
Electrical	*I31*	**-0.410**	0.90	0.28	Hand Tools	*I65*	**-0.454**	0.66	0.08
Electrical	*I32*	-0.154	0.79	0.17	Hand Tools	*I66*	**-0.436**	0.45	0.26
Electrical	*I33*	0.010	0.78	0.08	Hand Tools	*I67*	-0.297	0.69	0.24
Electrical	*I34*	**-0.717**	0.79	0.46	Hand Tools	*I68*	**-0.608**	0.56	0.31
Electrical	*I35*	**-0.727**	0.90	0.44	Hand Tools	*I69*	**-0.417**	0.84	0.31
Electrical	*I36*	**-0.683**	0.71	0.48	Hand Tools	*I70*	**-0.459**	0.91	0.29

Table 17 (continued)

Subsection	Item	Factor One	p_i	P.B.
Math/Stats	*I71*	**-0.597**	0.97	0.25
Math/Stats	*I72*	**-0.793**	0.97	0.35
Math/Stats	*I73*	**-0.563**	0.82	0.40
Math/Stats	*I74*	-0.131	0.82	0.13
Math/Stats	*I75*	**-0.605**	0.82	0.45
Math/Stats	*I76*	**-0.610**	0.86	0.43
Math/Stats	*I77*	**-0.558**	0.90	0.34
Math/Stats	*I78*	**-0.758**	0.94	0.42
Math/Stats	*I79*	**-0.321**	0.56	0.28
Math/Stats	*I80*	**-0.481**	0.96	0.25
Prob. Solve	*I81*	**-0.416**	0.76	0.34
Prob. Solve	*I82*	**-0.306**	0.86	0.23
Prob. Solve	*I83*	-0.116	0.37	0.13
Prob. Solve	*I84*	**-0.566**	0.84	0.40
Prob. Solve	*I85*	**-0.623**	0.90	0.41
Prob. Solve	*I86*	**-0.355**	0.77	0.28
Prob. Solve	*I87*	**-0.355**	0.84	0.22
Prob. Solve	*I88*	**-0.583**	0.98	0.24
Prob. Solve	*I89*	**-0.469**	0.89	0.31
Prob. Solve	*I90*	**-0.614**	0.95	0.33
Prob. Solve	*I91*	**-0.504**	0.89	0.34
Prob. Solve	*I92*	**-0.674**	0.79	0.50
Prob. Solve	*I93*	**-0.515**	0.94	0.30
Prob. Solve	*I94*	**-0.696**	0.86	0.47
Prob. Solve	*I95*	-0.179	0.80	0.16
Prob. Solve	*I96*	**-0.411**	0.79	0.33
Prob. Solve	*I97*	-0.157	0.84	0.14
Prob. Solve	*I98*	**-0.547**	0.96	0.27
Prob. Solve	*I99*	**-0.511**	0.92	0.30
Prob. Solve	*I100*	**-0.644**	0.95	0.34
Prob. Solve	*I101*	**-0.562**	0.83	0.42
Prob. Solve	*I102*	**-0.710**	0.96	0.31
Prob. Solve	*I103*	**-0.718**	0.95	0.35
Prob. Solve	*I104*	**-0.597**	0.88	0.38

Subsection	Item	Factor One	p_i	P.B.
Prob. Solve	*I105*	**-0.507**	0.64	0.44
Prob. Solve	*I106*	**-0.391**	0.87	0.27
Pneumatics	*I107*	**-0.441**	0.83	0.34
Pneumatics	*I108*	-0.175	0.75	0.16
Pneumatics	*I109*	**-0.419**	0.71	0.34
Pneumatics	*I110*	-0.065	0.68	0.11
Pneumatics	*I111*	**-0.316**	0.68	0.25
Pneumatics	*I112*	**-0.613**	0.89	0.42
Pneumatics	*I113*	**-0.362**	0.46	0.30
Pneumatics	*I114*	**-0.485**	0.42	0.41
Pneumatics	*I115*	**-0.583**	0.70	0.50
Pneumatics	*I116*	**-0.324**	0.60	0.31
Pneumatics	*I117*	**-0.722**	0.90	0.48
Pneumatics	*I118*	**-0.409**	0.53	0.36
Pneumatics	*I119*	**-0.427**	0.68	0.36
Pneumatics	*I120*	-0.004	0.21	0.05
Pneumatics	*I121*	-0.281	0.78	0.24
Pneumatics	*I122*	-0.288	0.77	0.25
Pneumatics	*I123*	**-0.628**	0.89	0.42
Pneumatics	*I124*	**-0.372**	0.82	0.30
Pneumatics	*I125*	**-0.336**	0.92	0.22
Pneumatics	*I126*	**-0.591**	0.92	0.38
Pneumatics	*I127*	-0.293	0.89	0.20
PA 1	*I128*	**-0.524**	0.78	0.40
PA 2	*I129*	**-0.370**	0.57	0.34
PA 3	*I130*	**-0.335**	0.54	0.29
PA 4	*I131*	**-0.388**	0.63	0.37
PA 5	*I132*	**-0.428**	0.62	0.38
PA 6	*I133*	**-0.555**	0.36	0.43
PA 7	*I134*	**-0.443**	0.48	0.37

4.3.2.8 Form B Two-Factor Solution Mplus version 3.11 was also used to conduct an EFA on Form B combined with the 7 PA exercises with two factors extracted. The factor correlation matrix is shown in Table 18. The Promax factor loadings and the structure coefficients for the first two factors as well as item statistics and subsections are shown in Table 19.

Table 18. Form B Factor Correlations

1.000	**0.224**
0.224	**1.000**

Table 19. Form B Promax Factor Loadings for Two Factors with Structure Coefficients and Item Statistics

		Promax Loadings		Structure Coefficients			
Subsection	**Item**	**Factor One**	**Factor Two**	**Factor One**	**Factor Two**	p_i	**P.B.**
Mechanical	***I3***	**0.409**	0.005	**0.410**	0.097	0.71	0.32
Mechanical	***I4***	**0.359**	0.043	**0.369**	0.123	0.71	0.31
Mechanical	***I5***	**0.345**	-0.039	**0.336**	0.038	0.36	0.27
Mechanical	***I6***	**0.522**	0.227	**0.573**	**0.344**	0.94	0.33
Mechanical	***I7***	**0.460**	-0.034	**0.452**	0.069	0.45	0.35
Mechanical	***I8***	**0.547**	-0.037	**0.539**	0.086	0.67	0.42
Mechanical	***I9***	**0.368**	0.064	**0.382**	0.146	0.65	0.31
Mechanical	***I10***	0.298	-0.084	0.279	-0.017	0.70	0.23
Mechanical	***I11***	**0.622**	-0.028	**0.616**	0.111	0.90	0.37
Mechanical	***I12***	**0.321**	**0.400**	**0.411**	**0.472**	0.97	0.22
Mechanical	***I13***	0.116	0.140	0.147	0.166	0.51	0.17
Mechanical	***I14***	**0.627**	-0.040	**0.618**	0.100	0.74	0.47
Mechanical	***I15***	**0.434**	0.120	**0.461**	0.217	0.93	0.25
Mechanical	***I16***	**0.427**	0.019	**0.431**	0.115	0.83	0.29
Mechanical	***I17***	**0.446**	-0.208	**0.399**	-0.108	0.81	0.25
Mechanical	***I18***	**0.381**	0.070	**0.397**	0.155	0.61	0.34
Mechanical	***I19***	0.050	0.027	0.056	0.038	0.28	0.07
Mechanical	***I20***	**0.379**	-0.208	**0.332**	-0.123	0.47	0.24
Mechanical	***I21***	**0.463**	-0.070	**0.447**	0.034	0.80	0.31
Mechanical	***I22***	**0.548**	-0.092	**0.527**	0.031	0.90	0.33
Electrical	***I23***	**0.475**	-0.101	**0.452**	0.005	0.96	0.20
Electrical	***I24***	**0.651**	0.066	**0.666**	0.212	0.66	0.51
Electrical	***I25***	**0.599**	-0.097	**0.577**	0.037	0.61	0.45
Electrical	***I26***	**0.476**	-0.040	**0.467**	0.067	0.80	0.32
Electrical	***I27***	**0.385**	0.024	**0.390**	0.110	0.90	0.26
Electrical	***I28***	**0.367**	-0.228	**0.316**	-0.146	0.53	0.25
Electrical	***I29***	**0.754**	0.008	**0.756**	0.177	0.81	0.53
Electrical	***I30***	**0.571**	**0.310**	**0.640**	**0.438**	0.96	0.33
Electrical	***I31***	**0.425**	0.024	**0.430**	0.119	0.90	0.28
Electrical	***I32***	0.186	-0.047	0.175	-0.005	0.79	0.17
Electrical	***I33***	0.048	-0.138	0.017	-0.127	0.78	0.08
Electrical	***I34***	**0.798**	-0.147	**0.765**	0.032	0.79	0.46
Electrical	***I35***	**0.759**	-0.012	**0.756**	0.158	0.90	0.44
Electrical	***I36***	**0.734**	-0.063	**0.720**	0.101	0.71	0.48
Electrical	***I37***	**0.614**	0.039	**0.623**	0.177	0.57	0.47
Electrical	***I38***	**0.734**	-0.056	**0.721**	0.108	0.91	0.42
Electrical	***I39***	**0.516**	-0.008	**0.514**	0.108	0.90	0.32
Electrical	***I40***	**0.467**	0.059	**0.480**	0.164	0.68	0.38

Table 19 (continued)

Subsection	Item	Promax Loadings Factor One	Promax Loadings Factor Two	Structure Coefficients Factor One	Structure Coefficients Factor Two	p_i	P.B.
Electrical	***I41***	0.187	-0.115	0.161	-0.073	0.54	0.17
Electrical	***I42***	**0.433**	-0.166	**0.396**	-0.069	0.56	0.32
Electrical	***I43***	**0.675**	0.016	**0.679**	0.167	0.88	0.45
Electrical	***I44***	**0.522**	-0.066	**0.507**	0.051	0.80	0.35
Electrical	***I45***	**0.327**	-0.085	**0.308**	-0.012	0.49	0.23
Electrical	***I46***	**0.526**	0.053	**0.538**	0.171	0.90	0.34
Electrical	***I47***	**0.710**	0.025	**0.716**	0.184	0.76	0.53
Electrical	***I48***	**0.593**	-0.017	**0.589**	0.116	0.73	0.44
Electrical	***I49***	0.207	**0.702**	**0.364**	**0.748**	0.37	0.37
Electrical	***I50***	**0.703**	0.201	**0.748**	**0.358**	0.90	0.51
Electrical	***I51***	**0.408**	-0.148	**0.375**	-0.057	0.84	0.27
Electrical	***I52***	0.136	**0.802**	**0.316**	**0.832**	0.46	0.38
Hydraulics	***I53***	0.021	**0.727**	0.184	**0.732**	0.51	0.27
Hydraulics	***I54***	0.023	**0.837**	0.210	**0.842**	0.87	0.26
Hydraulics	***I55***	0.137	**0.708**	0.296	**0.739**	0.57	0.36
Hydraulics	***I56***	0.203	-0.062	0.189	-0.017	0.90	0.11
Hydraulics	***I57***	0.005	**0.670**	0.155	**0.671**	0.45	0.24
Hydraulics	***I58***	-0.200	**1.029**	0.030	**0.984**	0.64	0.22
Hydraulics	***I59***	**0.540**	0.019	**0.544**	0.140	0.87	0.35
Hydraulics	***I60***	**0.530**	0.046	**0.540**	0.165	0.90	0.35
Hydraulics	***I61***	**0.615**	0.028	**0.621**	0.166	0.96	0.30
Hydraulics	***I62***	-0.032	**0.996**	0.191	**0.989**	0.64	0.32
Hand Tools	***I63***	0.208	**0.650**	**0.354**	**0.697**	0.88	0.32
Hand Tools	***I64***	0.158	**0.591**	0.290	**0.626**	0.26	0.31
Hand Tools	***I65***	**-0.360**	**0.941**	-0.149	**0.860**	0.66	0.08
Hand Tools	***I66***	0.006	**0.741**	0.172	**0.742**	0.45	0.26
Hand Tools	***I67***	**0.312**	0.018	**0.316**	0.088	0.69	0.24
Hand Tools	***I68***	0.021	**0.856**	0.213	**0.861**	0.56	0.31
Hand Tools	***I69***	**0.558**	-0.248	**0.502**	-0.123	0.84	0.31
Hand Tools	***I70***	**0.474**	0.035	**0.482**	0.141	0.91	0.29
Math/Stats	***I71***	**0.593**	0.072	**0.609**	0.205	0.97	0.25
Math/Stats	***I72***	**0.696**	**0.324**	**0.769**	**0.480**	0.97	0.35
Math/Stats	***I73***	**0.586**	0.024	**0.591**	0.155	0.82	0.40
Math/Stats	***I74***	0.165	-0.051	0.154	-0.014	0.82	0.13
Math/Stats	***I75***	**0.672**	-0.068	**0.657**	0.083	0.82	0.45
Math/Stats	***I76***	**0.638**	0.015	**0.641**	0.158	0.86	0.43
Math/Stats	***I77***	**0.536**	0.130	**0.565**	0.250	0.90	0.34
Math/Stats	***I78***	**0.701**	0.228	**0.752**	**0.385**	0.94	0.42
Math/Stats	***I79***	**0.331**	0.040	**0.340**	0.114	0.56	0.28
Math/Stats	***I80***	**0.567**	-0.088	**0.547**	0.039	0.96	0.25
Problem Solving	***I81***	**0.466**	-0.044	**0.456**	0.060	0.76	0.34
Problem Solving	***I82***	**0.366**	-0.091	**0.346**	-0.009	0.86	0.23
Problem Solving	***I83***	0.174	-0.097	0.152	-0.058	0.37	0.13

Table 19 (continued)

Subsection	Item	Promax Loadings Factor One	Promax Loadings Factor Two	Structure Coefficients Factor One	Structure Coefficients Factor Two	p_i	P.B.
Problem Solving	***I84***	**0.519**	0.190	**0.562**	**0.306**	0.84	0.40
Problem Solving	***I85***	**0.677**	-0.041	**0.668**	0.111	0.90	0.41
Problem Solving	***I86***	**0.370**	0.039	**0.379**	0.122	0.77	0.28
Problem Solving	***I87***	**0.329**	0.120	**0.356**	0.194	0.84	0.22
Problem Solving	***I88***	**0.563**	0.117	**0.589**	0.243	0.98	0.24
Problem Solving	***I89***	**0.463**	0.087	**0.482**	0.191	0.89	0.31
Problem Solving	***I90***	**0.548**	0.232	**0.600**	**0.355**	0.95	0.33
Problem Solving	***I91***	**0.581**	-0.117	**0.555**	0.013	0.89	0.34
Problem Solving	***I92***	**0.718**	-0.016	**0.714**	0.145	0.79	0.50
Problem Solving	***I93***	**0.506**	0.081	**0.524**	0.194	0.94	0.30
Problem Solving	***I94***	**0.700**	0.071	**0.716**	0.228	0.86	0.47
Problem Solving	***I95***	0.177	0.036	0.185	0.076	0.80	0.16
Problem Solving	***I96***	**0.444**	-0.006	**0.443**	0.093	0.79	0.33
Problem Solving	***I97***	0.180	-0.024	0.175	0.016	0.84	0.14
Problem Solving	***I98***	**0.488**	0.225	**0.538**	**0.334**	0.96	0.27
Problem Solving	***I99***	**0.385**	**0.357**	**0.465**	**0.443**	0.92	0.30
Problem Solving	***I100***	**0.560**	0.292	**0.625**	**0.417**	0.95	0.34
Problem Solving	***I101***	**0.569**	0.066	**0.584**	0.193	0.83	0.42
Problem Solving	***I102***	**0.747**	-0.028	**0.741**	0.139	0.96	0.31
Problem Solving	***I103***	**0.754**	-0.020	**0.750**	0.149	0.95	0.35
Problem Solving	***I104***	**0.618**	0.032	**0.625**	0.170	0.88	0.38
Problem Solving	***I105***	**0.542**	0.001	**0.542**	0.122	0.64	0.44
Problem Solving	***I106***	**0.413**	0.016	**0.417**	0.109	0.87	0.27
Pneumatics	***I107***	**0.506**	-0.079	**0.488**	0.034	0.83	0.34
Pneumatics	***I108***	0.213	-0.049	0.202	-0.001	0.75	0.16
Pneumatics	***I109***	**0.399**	0.112	**0.424**	0.201	0.71	0.34
Pneumatics	***I110***	0.142	-0.161	0.106	-0.129	0.68	0.11
Pneumatics	***I111***	**0.316**	0.044	**0.326**	0.115	0.68	0.25
Pneumatics	***I112***	**0.664**	-0.046	**0.654**	0.103	0.89	0.42
Pneumatics	***I113***	**0.377**	0.025	**0.383**	0.109	0.46	0.30
Pneumatics	***I114***	**0.554**	-0.059	**0.541**	0.065	0.42	0.41
Pneumatics	***I115***	**0.639**	-0.043	**0.629**	0.100	0.70	0.50
Pneumatics	***I116***	**0.369**	-0.060	**0.356**	0.023	0.60	0.31
Pneumatics	***I117***	**0.759**	-0.006	**0.758**	0.164	0.90	0.48
Pneumatics	***I118***	**0.508**	-0.136	**0.478**	-0.022	0.53	0.36
Pneumatics	***I119***	**0.437**	0.045	**0.447**	0.143	0.68	0.36
Pneumatics	***I120***	0.046	-0.089	0.026	-0.079	0.21	0.05
Pneumatics	***I121***	**0.341**	-0.098	**0.319**	-0.022	0.78	0.24
Pneumatics	***I122***	**0.309**	-0.001	**0.309**	0.068	0.77	0.25
Pneumatics	***I123***	**0.629**	0.081	**0.647**	0.222	0.89	0.42
Pneumatics	***I124***	**0.359**	0.077	**0.376**	0.157	0.82	0.30
Pneumatics	***I125***	**0.378**	-0.048	**0.367**	0.037	0.92	0.22
Pneumatics	***I126***	**0.663**	-0.111	**0.638**	0.038	0.92	0.38

Table 19 (continued)

Subsection	Item	Promax Loadings Factor One	Promax Loadings Factor Two	Structure Coefficients Factor One	Structure Coefficients Factor Two	p_i	P.B.
Pneumatics	***I127***	0.266	0.108	0.290	0.168	0.89	0.20
PA 1	***I128***	**0.565**	-0.011	**0.563**	0.116	0.78	0.40
PA 2	***I129***	**0.463**	-0.147	**0.430**	-0.043	0.57	0.34
PA 3	***I130***	0.298	0.131	**0.327**	0.198	0.54	0.29
PA 4	***I131***	**0.463**	-0.106	**0.439**	-0.002	0.63	0.37
PA 5	***I132***	**0.498**	-0.089	**0.478**	0.023	0.62	0.38
PA 6	***I133***	**0.684**	-0.201	**0.639**	-0.048	0.36	0.43
PA 7	***I134***	**0.500**	-0.058	**0.487**	0.054	0.48	0.37

4.3.2.9 Form B Two-Factor Varimax Solution For Form B, the correlation between the two factors was .224 making it difficult to justify using an oblique Promax rotation for interpretation. The low correlation suggested that an examination of the orthogonal, Varimax rotation as opposed to the oblique, Promax rotation was appropriate. The Varimax rotation produces a factor structure where the factors are uncorrelated, allowing each factor to represent a distinct construct. The resulting Varimax factor loadings for the first two factors as well as item statistics and subsections are presented in Table 20.

Table 20. Form B Varimax Factor Loadings for Two Factors with Item Statistics

Subsection	Item	Factor One (Varimax Rotated Loadings)	Factor Two (Varimax Rotated Loadings)	p_i	P.B.
Mechanical	*I3*	**0.402**	0.081	0.71	0.32
Mechanical	*I4*	**0.354**	0.110	0.71	0.31
Mechanical	*I5*	**0.338**	0.026	0.36	0.27
Mechanical	*I6*	**0.522**	**0.325**	0.94	0.33
Mechanical	*I7*	**0.451**	0.052	0.45	0.35
Mechanical	*I8*	**0.536**	0.065	0.67	0.42
Mechanical	*I9*	**0.364**	0.133	0.65	0.31
Mechanical	*I10*	0.290	-0.028	0.70	0.23
Mechanical	*I11*	**0.610**	0.088	0.90	0.37
Mechanical	*I12*	**0.330**	**0.460**	0.97	0.22
Mechanical	*I13*	0.119	0.161	0.51	0.17
Mechanical	*I14*	**0.614**	0.077	0.74	0.47
Mechanical	*I15*	**0.431**	0.201	0.93	0.25
Mechanical	*I16*	**0.420**	0.099	0.83	0.29
Mechanical	*I17*	**0.430**	-0.124	0.81	0.25
Mechanical	*I18*	**0.376**	0.141	0.61	0.34
Mechanical	*I19*	0.050	0.036	0.28	0.07
Mechanical	*I20*	**0.365**	-0.137	0.47	0.24
Mechanical	*I21*	**0.452**	0.017	0.80	0.31
Mechanical	*I22*	**0.535**	0.011	0.90	0.33
Electrical	*I23*	**0.463**	-0.012	0.96	0.20
Electrical	*I24*	**0.642**	0.188	0.66	0.51
Electrical	*I25*	**0.585**	0.015	0.61	0.45
Electrical	*I26*	**0.466**	0.049	0.80	0.32
Electrical	*I27*	**0.379**	0.096	0.90	0.26
Electrical	*I28*	**0.352**	-0.159	0.53	0.25
Electrical	*I29*	**0.741**	0.149	0.81	0.53
Electrical	*I30*	**0.573**	**0.416**	0.96	0.33
Electrical	*I31*	**0.419**	0.104	0.90	0.28
Electrical	*I32*	0.181	-0.013	0.79	0.17
Electrical	*I33*	0.042	-0.129	0.78	0.08
Electrical	*I34*	**0.778**	0.002	0.79	0.46
Electrical	*I35*	**0.745**	0.130	0.90	0.44
Electrical	*I36*	**0.718**	0.074	0.71	0.48
Electrical	*I37*	**0.605**	0.154	0.57	0.47
Electrical	*I38*	**0.719**	0.081	0.91	0.42
Electrical	*I39*	**0.507**	0.088	0.90	0.32
Electrical	*I40*	**0.461**	0.146	0.68	0.38
Electrical	*I41*	0.179	-0.080	0.54	0.17

Table 20 (continued).

Subsection	Item	Varimax Rotated Loadings Factor One	 Factor Two	p_i	P.B.
Electrical	***I42***	**0.419**	-0.085	0.56	0.32
Electrical	***I43***	**0.664**	0.142	0.88	0.45
Electrical	***I44***	**0.511**	0.032	0.80	0.35
Electrical	***I45***	**0.319**	-0.024	0.49	0.23
Electrical	***I46***	**0.519**	0.151	0.90	0.34
Electrical	***I47***	**0.699**	0.158	0.76	0.53
Electrical	***I48***	**0.582**	0.094	0.73	0.44
Electrical	***I49***	0.230	**0.740**	0.37	0.37
Electrical	***I50***	**0.699**	**0.332**	0.90	0.51
Electrical	***I51***	**0.395**	-0.071	0.84	0.27
Electrical	***I52***	0.164	**0.827**	0.46	0.38
Hydraulics	***I53***	0.048	**0.730**	0.51	0.27
Hydraulics	***I54***	0.054	**0.841**	0.87	0.26
Hydraulics	***I55***	0.161	**0.733**	0.57	0.36
Hydraulics	***I56***	0.197	-0.024	0.90	0.11
Hydraulics	***I57***	0.030	**0.670**	0.45	0.24
Hydraulics	***I58***	-0.157	**0.991**	0.64	0.22
Hydraulics	***I59***	**0.532**	0.120	0.87	0.35
Hydraulics	***I60***	**0.522**	0.145	0.90	0.35
Hydraulics	***I61***	**0.605**	0.143	0.96	0.30
Hydraulics	***I62***	0.006	**0.990**	0.64	0.32
Hand Tools	***I63***	0.229	**0.688**	0.88	0.32
Hand Tools	***I64***	0.178	**0.620**	0.26	0.31
Hand Tools	***I65***	**-0.318**	**0.873**	0.66	0.08
Hand Tools	***I66***	0.034	**0.741**	0.45	0.26
Hand Tools	***I67***	**0.308**	0.076	0.69	0.24
Hand Tools	***I68***	0.053	**0.859**	0.56	0.31
Hand Tools	***I69***	**0.539**	-0.143	0.84	0.31
Hand Tools	***I70***	**0.467**	0.124	0.91	0.29
Math/Stats	***I71***	**0.585**	0.182	0.97	0.25
Math/Stats	***I72***	**0.696**	**0.453**	0.97	0.35
Math/Stats	***I73***	**0.577**	0.134	0.82	0.40
Math/Stats	***I74***	0.160	-0.020	0.82	0.13
Math/Stats	***I75***	**0.657**	0.058	0.82	0.45
Math/Stats	***I76***	**0.627**	0.134	0.86	0.43
Math/Stats	***I77***	**0.532**	0.230	0.90	0.34
Math/Stats	***I78***	**0.697**	**0.359**	0.94	0.42
Math/Stats	***I79***	**0.327**	0.102	0.56	0.28
Math/Stats	***I80***	**0.554**	0.018	0.96	0.25

Table 20 (continued).

Subsection	Item	Factor One	Factor Two	p_i	P.B.
		Varimax Rotated Loadings			
Problem Solving	***I81***	**0.456**	0.043	0.76	0.34
Problem Solving	***I82***	**0.356**	-0.023	0.86	0.23
Problem Solving	***I83***	0.168	-0.064	0.37	0.13
Problem Solving	***I84***	**0.517**	0.287	0.84	0.40
Problem Solving	***I85***	**0.664**	0.085	0.90	0.41
Problem Solving	***I86***	**0.365**	0.108	0.77	0.28
Problem Solving	***I87***	**0.327**	0.181	0.84	0.22
Problem Solving	***I88***	**0.557**	0.222	0.98	0.24
Problem Solving	***I89***	**0.458**	0.173	0.89	0.31
Problem Solving	***I90***	**0.547**	**0.334**	0.95	0.33
Problem Solving	***I91***	**0.567**	-0.008	0.89	0.34
Problem Solving	***I92***	**0.705**	0.118	0.79	0.50
Problem Solving	***I93***	**0.500**	0.176	0.94	0.30
Problem Solving	***I94***	**0.690**	0.201	0.86	0.47
Problem Solving	***I95***	0.175	0.069	0.80	0.16
Problem Solving	***I96***	**0.436**	0.077	0.79	0.33
Problem Solving	***I97***	0.176	0.010	0.84	0.14
Problem Solving	***I98***	**0.488**	**0.316**	0.96	0.27
Problem Solving	***I99***	**0.392**	**0.429**	0.92	0.30
Problem Solving	***I100***	**0.561**	**0.397**	0.95	0.34
Problem Solving	***I101***	**0.562**	0.172	0.83	0.42
Problem Solving	***I102***	**0.733**	0.112	0.96	0.31
Problem Solving	***I103***	**0.740**	0.121	0.95	0.35
Problem Solving	***I104***	**0.608**	0.147	0.88	0.38
Problem Solving	***I105***	**0.532**	0.102	0.64	0.44
Problem Solving	***I106***	**0.406**	0.093	0.87	0.27
Pneumatics	***I107***	**0.494**	0.015	0.83	0.34
Pneumatics	***I108***	0.208	-0.009	0.75	0.16
Pneumatics	***I109***	**0.396**	0.187	0.71	0.34
Pneumatics	***I110***	0.134	-0.134	0.68	0.11
Pneumatics	***I111***	**0.312**	0.103	0.68	0.25
Pneumatics	***I112***	**0.651**	0.078	0.89	0.42
Pneumatics	***I113***	**0.371**	0.095	0.46	0.30
Pneumatics	***I114***	**0.542**	0.045	0.42	0.41
Pneumatics	***I115***	**0.626**	0.077	0.70	0.50
Pneumatics	***I116***	**0.360**	0.009	0.60	0.31
Pneumatics	***I117***	**0.746**	0.136	0.90	0.48
Pneumatics	***I118***	**0.494**	-0.041	0.53	0.36
Pneumatics	***I119***	**0.431**	0.127	0.68	0.36

Table 20 (continued).

Subsection	Item	Varimax Rotated Loadings Factor One	Factor Two	p_i	P.B.
Pneumatics	***I120***	0.042	-0.080	0.21	0.05
Pneumatics	***I121***	**0.331**	-0.034	0.78	0.24
Pneumatics	***I122***	**0.303**	0.057	0.77	0.25
Pneumatics	***I123***	**0.621**	0.198	0.89	0.42
Pneumatics	***I124***	**0.356**	0.145	0.82	0.30
Pneumatics	***I125***	**0.369**	0.023	0.92	0.22
Pneumatics	***I126***	**0.647**	0.013	0.92	0.38
Pneumatics	***I127***	0.265	0.157	0.89	0.20
PA 1	***I128***	**0.555**	0.095	0.78	0.40
PA 2	***I129***	**0.449**	-0.060	0.57	0.34
PA 3	***I130***	0.297	0.187	0.54	0.29
PA 4	***I131***	**0.450**	-0.020	0.63	0.37
PA 5	***I132***	**0.486**	0.004	0.62	0.38
PA 6	***I133***	**0.664**	-0.073	0.36	0.43
PA 7	***I134***	**0.489**	0.035	0.48	0.37

For the Varimax rotation, the items that loaded on factor two were located in 6 of the 7 subsections on the MC test. The items that loaded above .3 on factor two were items 6, 12, 30, 49, 50, 52, 53, 54, 55, 57, 58, 62, 63, 64, 65, 66, 68, 72, 78, 90, 98, 99 and 100. Half of the items that loaded on the second factor were located in the subsections of Hydraulics/Fluid Flow (items 53, 54, 55, 57, 58, 62) and Hand, Measuring Tool and Equipment (items 63, 64, 65, 66, and 68). Of those items, 65 had a point biserial correlation of .08 and three others (57, 58 and 66) had point biserial correlations below .30.

A review of item content was conducted to attempt to explain the apparent presence of a second factor on Form B. A discussion follows regarding items that had both a factor two loading larger than .3, and the factor one loading, and a comparison of these items to other items in their test section.

An attempt was made to specify each item's content (in parenthesis after each item number) using a one or two word description from the knowledge in skill areas that were used to develop the tests as shown in Table 21. In some cases, an appropriate one or two word description could not be obtained from the original knowledge and skill areas in Table 21 so a new description was created. These new one or two word descriptions are in *italics*.

Table 21. L2/L3 Production Technician Knowledge and Skills Areas

Average Ranked Importance	Average % of Items	Estimated No. of Questions	Area		
2	20.8	25	A. Mechanical		
			1. Troubleshoot	5. Flow paths	9. Pumps
			2. Repair/replace	6. Alignment	10. Valves
			3. Principles	7. Gear boxes	11. Assembly drawings
			4. Operations	8. Conveyors	
4	16.7	20	B. Electrical		
			1. Troubleshoot	5. Heaters	9. PLC systems
			2. Repair/replace	6. Motors	10. Electrical drawings
			3. Servos	7. Blowers/fans	
			4. Switches	8. AC/DC circuits	
7	5.8	10	C. Hydraulics (Fluid Flow)		
			1. Troubleshoot	4. Accumulators	7. Hydraulic prints
			2. Repair/replace	5. Control valves	
			3. Pumps	6. Hoses	
6	6.7	10	D. Hand/Measuring Tools		
			1. Wrenches	3. Voltmeters	5. Gauges
			2. Multimeters	4. Calipers	
5	8.3	10	E. Math/Statistics		
			1. X/Y axis	3. Decimals	5. Averages
			2. Percentages	4. Fractions	
1	24.2	29	F. Problem Solving		
			1. Cause & effect	3. Resolve issues	5. Pareto & pie charts
			2. Interpret data	4. Bar/line graphs	6. Flow diagrams
3	17.5	21	G. Pneumatics		
			1. Troubleshoot	4. Air filters	7. Valves
			2. Repair/replace	5. Hoses	8. Vacuum systems
			3. Air cylinders	6. Nozzles	9. Fittings/couplings
		125	Safety was found to be pervasive		

Note. From *Content validation report: Assessment selection & development [L2&L3 (Production Technician)]* by R.T. Ramsay, 2000, p. 11.

Based on a review of item content, all of the items in the Mechanical section appear to represent an applied mechanical knowledge dimension. Item 12 (conveyors) from the Mechanical test section had a .330 loading on factor one and a .460 loading on factor two. In terms of content, item 12 (conveyors) appears to be most similar to item 13 (conveyors) as the specific content for both items refers to conveyor belt issues. However, item 13 (conveyors) did not load above .3 on either factor. The non-loading of item 13 (conveyors) is consistent with its low (.17) point biserial correlation, while the loadings for item 12 (conveyors) may be a reflection of its high item difficulty index (.97). In fact, all the items in the Mechanical test section loaded on factor one with the exception of items 10 (principles), 13 (conveyors), and 19 (principles) which had low point biserial correlations (.23, .17 and .07 respectively). The low point biserial correlations indicate ineffective items for employee selection purposes.

In terms of content, all of the items in the Electrical test section reflect an applied electrical knowledge dimension. In terms of specific item content, item 49 (PLC systems), and item 52 (PLC systems) which both loaded highly on factor two, are most similar to items 50 (PLC systems), and 51 (PLC systems) which both loaded on factor one. An analysis of the content of these items reveals no apparent differences with each other or the other items in the test section which would explain the presence of a second factor.

An analysis of the content of the Hydraulics/Fluid Flow test section reveals that all of the items reflect an applied hydraulics knowledge dimension. Items 53 (repair/replace), items 54 (repair/replace), 55 (repair/replace), 57 (repair/replace) and 58 (repair/replace) and 62 (repair/replace), which loaded on factor two are most similar to the content of items 60 (repair/replace) and 61 (repair/replace) which loaded on factor one. Item 56 (pumps) did not load on either factor, while item 59 (*principles*) loaded on factor one. An analysis of the content of these items reveals no apparent differences with each other or the other items in the test section which would explain the presence of a second factor.

Based on a review of item content, all of the items in the Hand, Measuring Tools and Equipment test section appear to represent an applied knowledge of mechanical tools and equipment. More specifically, items 63 (gauges), 65 (gauges), and 68 (gauges) which loaded on factor two, represent the only measurement content of the section. However, item 65 had a very low point biserial correlation of .08. Items 64 (*mechanical principles*) which also loaded on factor two is most similar to item 70 (*mechanical principles*) which loaded on factor one. Item 66 (*troubleshoot*) which loaded on factor two, was most similar to item 67 (*troubleshoot*) which loaded on factor one. An analysis of the content of these items reveals no apparent differences with each other or the other items in the test section which would explain the presence of factor two.

An analysis of the content of the Problem solving test section reveals that all of the items reflect an applied problem solving/troubleshooting dimension. Item 99 (flow diagrams), 100 (flow diagrams), and 101 (flow diagrams) all refer to the same flow chart diagram. However, item 101(flow diagrams) loads solely on factor one, while items 99 (flow diagrams) and 100 (flow diagrams) crossload on both factor one and factor two. Of these three items only item 99 (flow diagrams) had both a Varimax rotated loading greater than .3 and a larger loading on factor two than on factor one. An analysis of the content of these three items reveals no apparent differences with the other items in the test section that would explain the factor two loadings. Items 83 (flow diagrams), 95 (cause and effect), and 97 (resolve issues) did not load on either factor, which is consistent with their corresponding low point biserial correlations (.13, .16, and .14 respectively).

4.3.2.10 Form B Crossloading Items A crossloading item loads at .3 or higher on two or more factors. Of the fifteen items that had both a Varimax rotated loading greater than .3 and a loading larger loading on factor two than on factor one, three could be considered crossloading items (items 12 (conveyors), 65 (gauges) and 99 (flow diagrams). Additionally, although they did not have a loading larger loading on factor two than on factor one, items 6 (pumps) , 30 (switches), 50 (PLC systems), 72 (*word problems*), 78 (decimals), 90 (interpret data), 98 (bar/line graph), and 100 (flow diagrams) also crossloaded on both factor one and factor two. These crossloadings may indicate items that are not working as expected.

4.3.2.11 Form B Second Factor Explanation The content review revealed no apparent difference in item content or construct on Form B between those fourteen items that loaded above .3 on factor two and the other items in their corresponding test section that would account for the presence of a second factor. Many of the items that loaded on factor two reflected the same dimension and featured similar content to items that loaded on factor one, or in some cases, neither factor.

An additional examination of the Form B Varimax factor loadings was conducted using a cutoff value of .4 instead of .3 in order clarify the interpretation of the factor structure. The results of this analysis produced additional support for the appropriateness of a one-factor solution.

One possibility for the appearance of a possible second factor could be related to the smaller sample size available for Form B (n=324) compared to Form A (n=432). Smaller sample sizes are more likely to result in items that are mis-classified on the wrong factor. EFA is a large sample procedure and generalizable or replicable results are unlikely if the sample size is too small. The smaller sample size coupled with sampling error from domain sampling may account for the presence of an apparent second factor on Form B. In fact, the RMSEA value of .070 did indicate moderate fit of the model for the one-factor solution for Form B. This taken together with the fact that the analysis of item content did not reveal an explanation for a second factor, indicated that the one-factor solution was most appropriate for Form B.

4.4 RESEARCH QUESTION 4 - WHAT IS THE RELATIONSHIP BETWEEN SCORES ON THE MC JOB KNOWLEDGE TESTS AND TOTAL SCORE ON THE PA EXERCISES?

Correlations were calculated for (a) MC total score with PA total score and (b) each MC test section score with PA total score. The results are shown in Table 22.

Table 22. MC Subsection Score with PA Total Score Correlations

Form A Subsection Correlations with PA Total

		Mech.	Elec.	Hydr.	Hand/ Tool	Math/ Stat.	Prob. Solve.	Pneum.	Form A Total
PA Total	Pearson Correl.	.527**	.518**	.484**	.323**	.368**	.520**	.481**	.627**
	Sig. (2-tailed)	0	0	0	0	0	0	0	0
	N	432	432	432	432	432	432	432	432

Form B Subsection Correlations with PA Total

		Mech.	Elec.	Hydr.	Hand/ Tool	Math/ Stat.	Prob. Solve.	Pneum.	Form B Total
PA Total	Pearson Correl.	.524**	.559**	.153**	.162**	.417**	.478**	.591**	.612**
	Sig. (2-tailed)	0	0	0.006	0.004	0	0	0	0
	N	324	324	324	324	324	324	324	324

** Correlation is significant at the 0.01 level (2-tailed).

The correlations for MC total score with PA total score were very similar for both Form A and Form B (.627 with .612 respectively). The MC subsections all showed significant positive correlations with PA total score for both test forms. Furthermore, the correlations for each MC subsection with PA total score were of similar magnitude for both forms with the exception of Hydraulics (.484 versus .153 for Form A and Form B respectively) and Hand/Measuring Tools (.323 versus .162 for Form A and Form B respectively). The MC test subsection with PA total score correlations differed somewhat in magnitude between Form A and Form B which was likely due to sampling error. However, as expected the two MC test forms had high significant positive correlations with PA total score.

The correlation for MC total score with PA total score for Form A (.627) was compared to the content experts' mean relevance rating (0.36) of the total MC test with the total PA test. The rather low relevance rating of the MC test with the PA test was surprising especially in light of the high positive correlation between the two types of items.

4.4.1 4a. Based on a regression analysis, what is the relationship between the predictor variable MC test score and the dependent variable PA test score?

A regression analysis was conducted using for both Form A and Form B using MC test scores as the independent variable and the PA test as the dependent variable.

4.4.1.1 Form A In order to examine linearity, a scatterplot of MC test scores (the independent variable) with PA total scores (the dependent variable) along with the computed regression line was plotted. The resulting plot for Form A with PA total score is presented in Figure 8. The slope for the regression line was .081 and the intercept was -3.533.

Scatterplot of Form A Total Score with PA Total Score

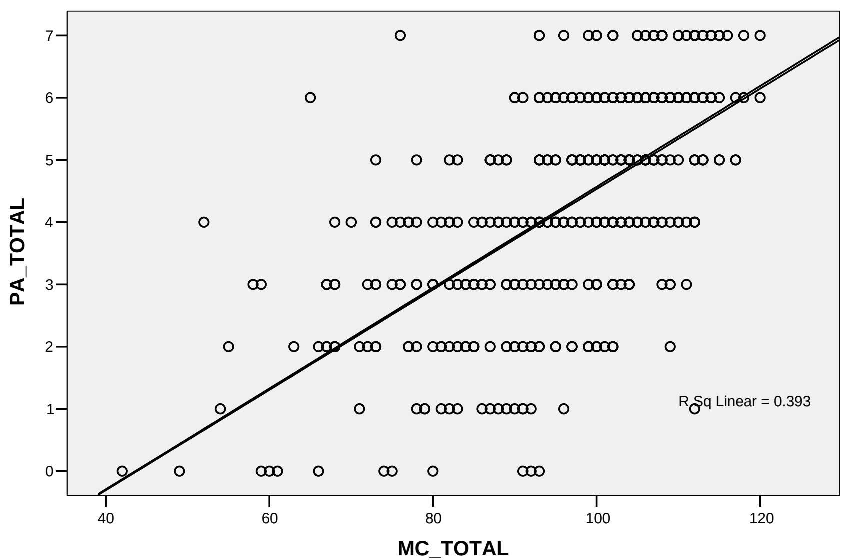

Figure 9. Scatterplot of Form A Total Score with PA Total Score

The *F* statistic which tests the hypothesis that the slope of the regression line is other than zero was also calculated. The ANOVA table is presented in Table 23.

Table 23. Analysis of Variance for Form A and PA Total Score

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	539.160	1	539.160	278.614	.000[a]
	Residual	832.115	430	1.935		
	Total	1371.275	431			

a. Predictors: (Constant), MC_TOTAL

b. Dependent Variable: PA_TOTAL

The *F* value was 278.614 and had an observed significance level less than .0005 indicating that the slope of the regression line is significantly different from zero.

Another measure of goodness of fit for the linear model, the R-square value, was also calculated. The R-square value, the Adjusted R-square, and the Standard Error of the Estimate are presented in Table 24. The R-square value was .393 which indicates the amount of the variability accounted for given the variables specified in the model. The R-square value for the Form A and PA total score analysis indicates that just under half of all of the variability in PA total score is accounted for with MC total score.

Table 24. Goodness of Fit Statistics for Form A

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.627[a]	.393	.392	1.391

a. Predictors: (Constant), MC_TOTAL

b. Dependent Variable: PA_TOTAL

An examination of the residuals was conducted to find if there was evidence that the necessary assumptions were violated. Figure 9 indicates that the assumptions of linearity and equality of variance are satisfied as the residuals appear to be randomly distributed.

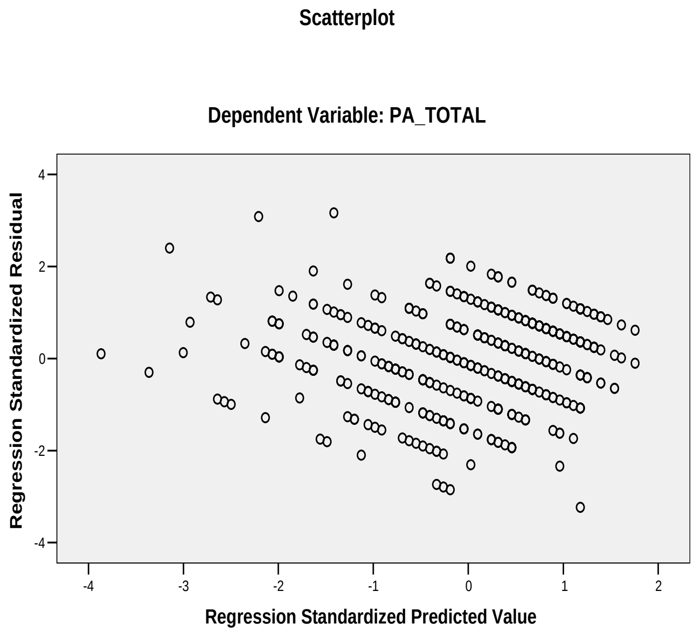

Figure 10. Form A Standardized Residuals Scatterplot

A histogram of the residuals was also produced to examine the assumption of normality. The histogram for Form A and PA total score is shown in Figure 11. The histogram of the residuals appears to be approximately normal.

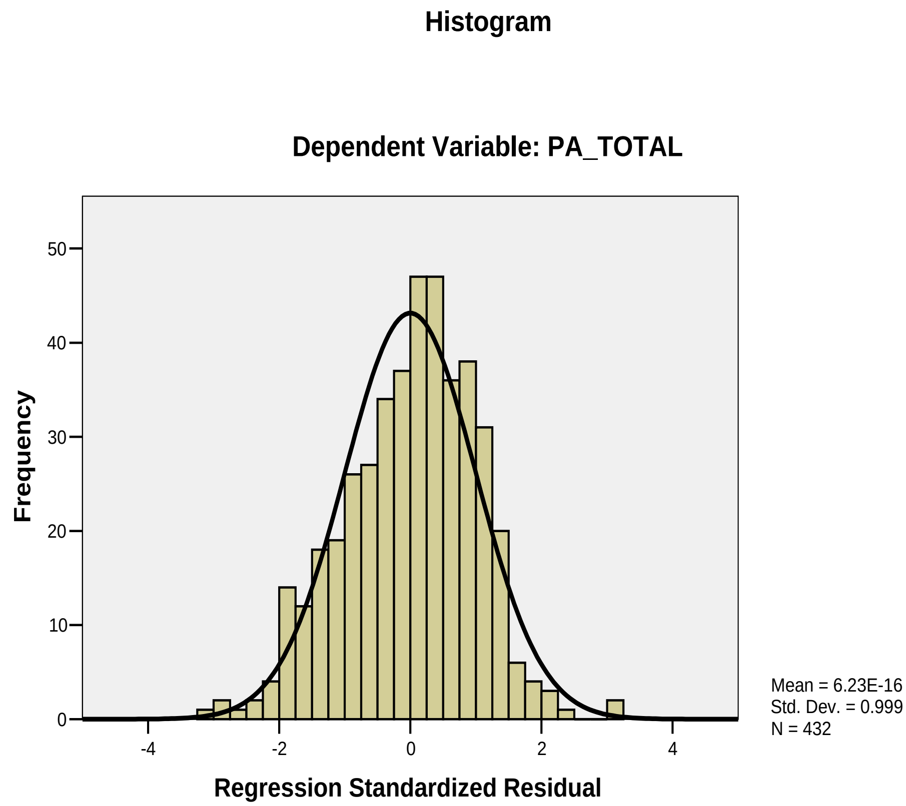

Figure 11. Histogram of Standardized Residuals for Form A

A cumulative probability plot of the residuals was also produced in order to examine the normality assumption. The Normal P-P of Regression Standardized Residuals is shown in Figure 12. This plot also indicates that the normality assumption is tenable.

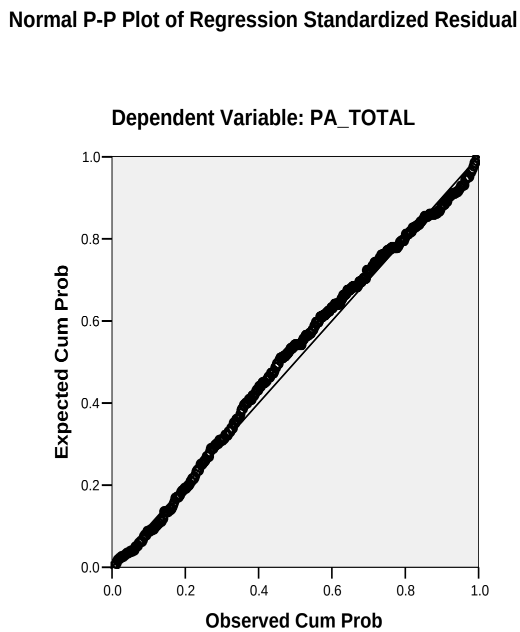

Figure 12. Normal Probability (P-P) Plot for Form A

4.4.1.2 Form B In order to examine linearity, a scatterplot of MC test scores (the independent variable) with PA total scores (the dependent variable) along with the computed regression line was plotted. The resulting plot for Form A with PA total score is presented in Figure 13. The slope for the regression line was .074 and the intercept was -3.051.

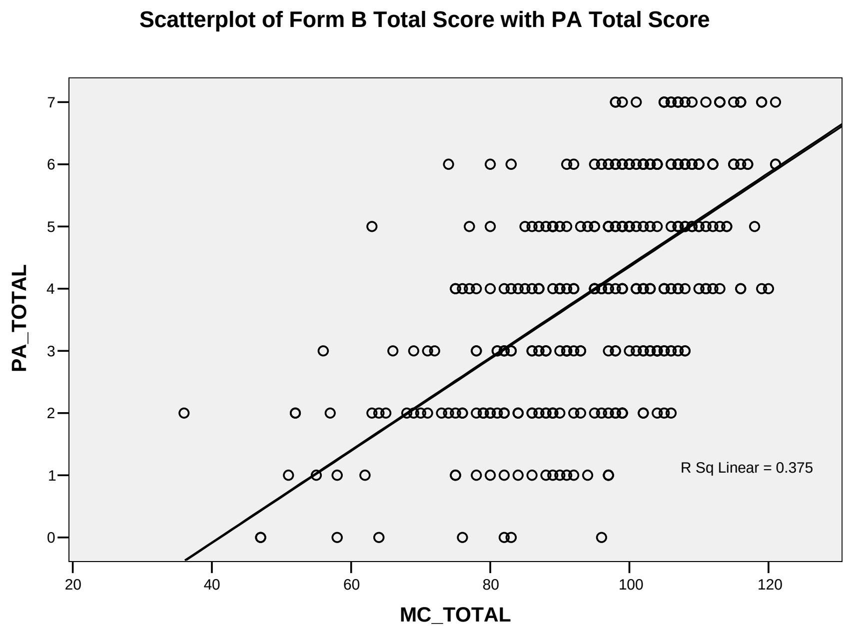

Figure 13. Scatterplot of Form B Total Score with PA Total Score

The *F* statistic which tests the hypothesis that the slope of the regression line is other than zero was also calculated. The ANOVA table is presented in Table 25.

Table 25. Analysis of Variance for Form B and PA Score

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	397.229	1	397.229	193.079	.000[a]
	Residual	662.463	322	2.057		
	Total	1059.691	323			

a. Predictors: (Constant), MC_TOTAL

b. Dependent Variable: PA_TOTAL

The *F* value was 193.079 and had an observed significance level less than .0005 indicating that the slope of the regression line is significantly different from zero.

Another measure of goodness of fit for the linear model, the R-square value, was also calculated. The R-square value, the Adjusted R-square, and the Standard Error of the Estimate are presented in Table 26. The R-square value was .375 which indicates the amount of the variability accounted for given the variables specified in the model. The R-square value for the Form A and PA total score analysis indicates that just under half of all of the variability in PA total score is accounted for with MC total score.

Table 26. Goodness of Fit Statistics for Form B

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.612[a]	.375	.373	1.434

a. Predictors: (Constant), MC_TOTAL

b. Dependent Variable: PA_TOTAL

An examination of the residuals was conducted to find if there was evidence that the necessary assumptions were violated. Figure 14 indicates that the assumptions of linearity and equality of variance are satisfied as the residuals appear to be randomly distributed.

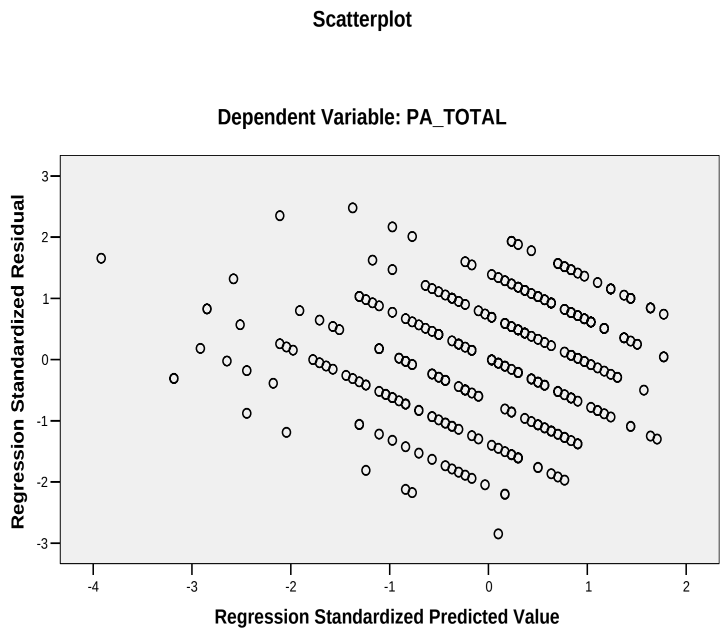

Figure 14. Form B Standardized Residuals Scatterplot

A histogram of the residuals was also produced to examine the assumption of normality. The histogram for Form B and PA total score is shown in Figure 15. The histogram of the residuals appears to be approximately normal.

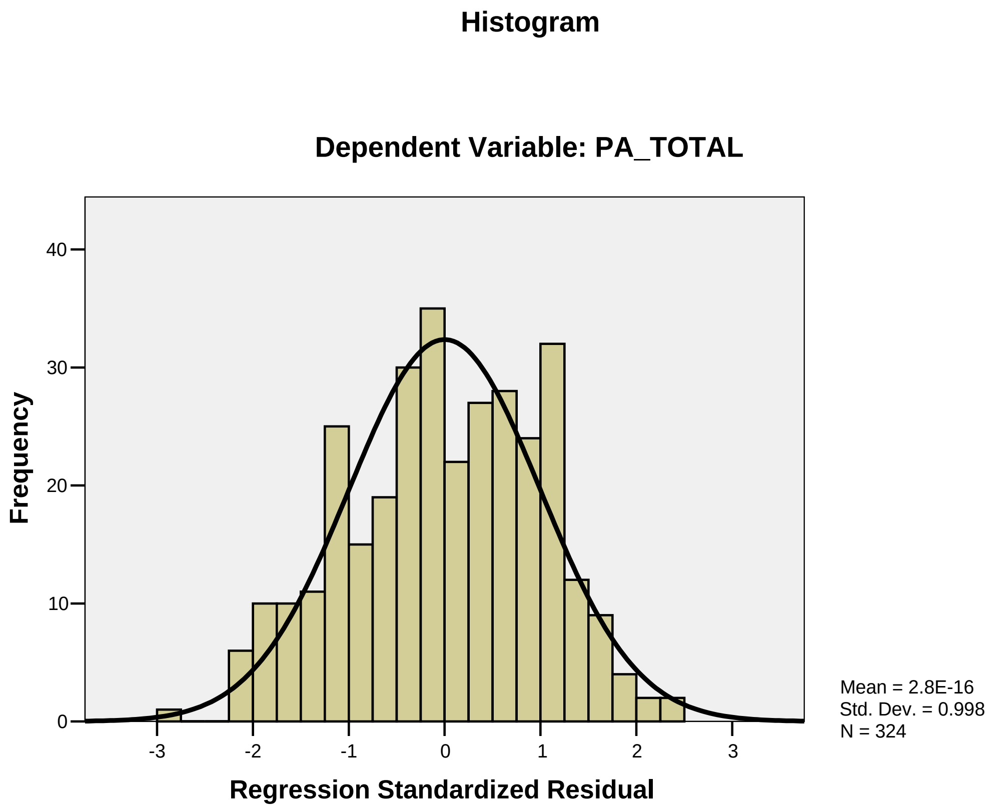

Figure 15. Histogram of Standardized Residuals for Form B

A cumulative probability plot of the residuals was also produced in order to examine the normality assumption. The Normal P-P of Regression Standardized Residuals is shown in Figure 16. This plot also indicates that the normality assumption is tenable.

Normal P-P Plot of Regression Standardized Residual

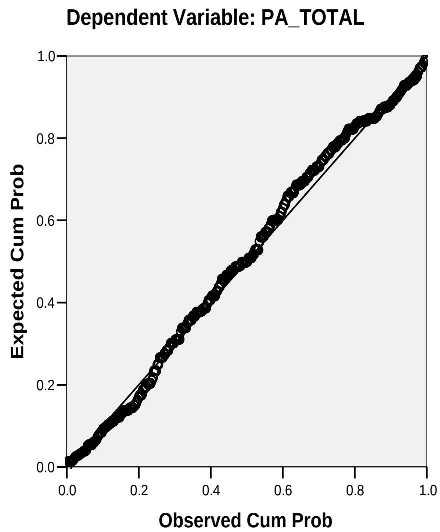

Figure 16. Normal Probability (P-P) Plot for Form B

The linear regression results for Form A and Form B were very similar in all aspects and there appeared to be no evidence of assumption violations.

4.4.2 4b. What is the decision consistency regarding those who score above and below the cut score on the MC test and the PA exercises?

Decisions for an examinee are consistent when the results of both the MC test and the PA test indicate that an examinee should be classified as passing. The percentage of consistent classifications from the MC test and the PA test was calculated by summing the percentage of examinees that passed both the MC test and the PA test with the percentage of examinees who failed both tests.

The percentages of passes and fails for Form A and Form B are shown in Figures 17 and 18 respectively. The decision consistency for Form A was 0.70 and the decision consistency for

Form B was 0.73. The probability of an inconsistent decision was 0.30 for Form A and 0.27 for Form B. Thus, both forms of the MC test resulted in similarly consistent decisions with the seven PA exercises.

Decisions Based on MC Test - Form A

Decisions Based on PA Test	Fail	Pass
Fail	0.47	0.05
Pass	0.25	0.23

Figure 17. Form A MC and PA Decision Consistency

Decisions Based on MC Test - Form B

Decisions Based on PA Test	Fail	Pass
Fail	0.54	0.05
Pass	0.22	0.19

Figure 18. Form B MC and PA Decision Consistency

The number of examinees who pass the MC test but then fail the PA test would likely be given the most influence by an employer weighing the option of using only the MC test for the L2/L3 selection procedure. For both forms, 5 percent passed the MC test and then failed the PA test, indicating that only of small percentage of false-positives would occur if only the MC test was used in the selection process. However, if the PA exercises are considered by the employer to be closer approximations of the L1/L2 Production Technician job than the MC tests, then the large percentage of false-negatives (25 percent for Form A and 22 percent for Form B) would likely be too high to justify using only the MC test to select L1/L3 Production Technicians.

4.5 RESEARCH QUESTION 5 - WHAT IS THE RELATIONSHIP BETWEEN EACH SUBTEST OF THE MC JOB KNOWLEDGE TEST AND EACH OF THE SEVEN PA EXERCISES?

Correlations between the PA exercises and the seven subtests on Form A and Form B are shown in Tables 27 and 28 respectively.

Table 27. Form A Subsections and PA Exercises Correlations

Correlations

		MECHANICAL	ELECTRICAL	HYDRAULICS	HAND_ TOOLS	MATH_ STATISTICS	PROBLEM_ SOLVING	PNEUMATICS	PA_1	PA_2	PA_3	PA_4	PA_5	PA_6	PA_7
MECHANICAL	Pearson Correlation	1	.598**	.571**	.484**	.517**	.632**	.465**	.295**	.298**	.257**	.265**	.349**	.292**	.254**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
ELECTRICAL	Pearson Correlation	.598**	1	.550**	.417**	.417**	.564**	.509**	.259**	.275**	.213**	.273**	.415**	.265**	.267**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
HYDRAULICS	Pearson Correlation	.571**	.550**	1	.507**	.337**	.573**	.463**	.304**	.243**	.195**	.316**	.348**	.251**	.207**
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
HAND_TOOLS	Pearson Correlation	.484**	.417**	.507**	1	.312**	.458**	.425**	.252**	.211**	.095*	.263**	.228**	.118*	.095*
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.000	.000	.047	.000	.000	.014	.049
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
MATH_STATISTICS	Pearson Correlation	.517**	.417**	.337**	.312**	1	.636**	.382**	.188**	.211**	.192**	.213**	.228**	.217**	.151**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000	.002
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
PROBLEM_SOLVING	Pearson Correlation	.632**	.564**	.573**	.458**	.636**	1	.529**	.308**	.302**	.254**	.292**	.336**	.272**	.226**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
PNEUMATICS	Pearson Correlation	.465**	.509**	.463**	.425**	.382**	.529**	1	.249**	.254**	.201**	.382**	.350**	.228**	.174**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
PA_1	Pearson Correlation	.295**	.259**	.304**	.252**	.188**	.308**	.249**	1	.157**	.124*	.201**	.167**	.221**	.149**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000		.001	.010	.000	.000	.000	.002
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
PA_2	Pearson Correlation	.298**	.275**	.243**	.211**	.211**	.302**	.254**	.157**	1	.144**	.250**	.306**	.193**	.072
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.001		.003	.000	.000	.000	.137
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
PA_3	Pearson Correlation	.257**	.213**	.195**	.095*	.192**	.254**	.201**	.124*	.144**	1	.258**	.156**	.137**	.085
	Sig. (2-tailed)	.000	.000	.000	.047	.000	.000	.000	.010	.003		.000	.001	.004	.078
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
PA_4	Pearson Correlation	.265**	.273**	.316**	.263**	.213**	.292**	.382**	.201**	.250**	.258**	1	.315**	.196**	.092
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000		.000	.000	.056
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
PA_5	Pearson Correlation	.349**	.415**	.348**	.228**	.228**	.336**	.350**	.167**	.306**	.156**	.315**	1	.142**	.149**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.000		.003	.002
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
PA_6	Pearson Correlation	.292**	.265**	.251**	.118*	.217**	.272**	.228**	.221**	.193**	.137**	.196**	.142**	1	.220**
	Sig. (2-tailed)	.000	.000	.000	.014	.000	.000	.000	.000	.000	.004	.000	.003		.000
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432
PA_7	Pearson Correlation	.254**	.267**	.207**	.095*	.151**	.226**	.174**	.149**	.072	.085	.092	.149**	.220**	1
	Sig. (2-tailed)	.000	.000	.000	.049	.002	.000	.000	.002	.137	.078	.056	.002	.000	
	N	432	432	432	432	432	432	432	432	432	432	432	432	432	432

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 28. Form B Subsections and PA Exercises Correlations

Correlations

		MECHANICAL	ELECTRICAL	HYDRAULICS	HAND_TOOL	MATH_ STATISTICS	PROBLEM_ SOLVING	PNEUMATICS	PA_1	PA_2	PA_3	PA_4	PA_5	PA_6	PA_7
MECHANICAL	Pearson Correlation	1	.619**	.234**	.267**	.452**	.552**	.582**	.326**	.307**	.182**	.309**	.286**	.331**	.245**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	.000	.000
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
ELECTRICAL	Pearson Correlation	.619**	1	.303**	.332**	.521**	.564**	.627**	.289**	.284**	.277**	.308**	.264**	.365**	.323**
	Sig. (2-tailed)	.000		.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
HYDRAULICS	Pearson Correlation	.234**	.303**	1	.740**	.202**	.261**	.242**	.120*	.050	.152**	.059	.080	.037	.084
	Sig. (2-tailed)	.000	.000		.000	.000	.000	.000	.031	.372	.006	.293	.153	.504	.131
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
HAND_TOOL	Pearson Correlation	.267**	.332**	.740**	1	.233**	.299**	.246**	.031	.033	.142*	.086	.124*	.051	.132*
	Sig. (2-tailed)	.000	.000	.000		.000	.000	.000	.580	.550	.011	.120	.025	.362	.017
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
MATH_STATISTICS	Pearson Correlation	.452**	.521**	.202**	.233**	1	.652**	.449**	.407**	.149**	.177**	.201**	.241**	.211**	.216**
	Sig. (2-tailed)	.000	.000	.000	.000		.000	.000	.000	.007	.001	.000	.000	.000	.000
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
PROBLEM_SOLVING	Pearson Correlation	.552**	.564**	.261**	.299**	.652**	1	.620**	.320**	.209**	.198**	.201**	.277**	.349**	.262**
	Sig. (2-tailed)	.000	.000	.000	.000	.000		.000	.000	.000	.000	.000	.000	.000	.000
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
PNEUMATICS	Pearson Correlation	.582**	.627**	.242**	.246**	.449**	.620**	1	.322**	.312**	.154**	.380**	.344**	.428**	.297**
	Sig. (2-tailed)	.000	.000	.000	.000	.000	.000		.000	.000	.005	.000	.000	.000	.000
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
PA_1	Pearson Correlation	.326**	.289**	.120*	.031	.407**	.320**	.322**	1	.148**	.099	.178**	.205**	.210**	.188**
	Sig. (2-tailed)	.000	.000	.031	.580	.000	.000	.000		.008	.074	.001	.000	.000	.001
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
PA_2	Pearson Correlation	.307**	.284**	.050	.033	.149**	.209**	.312**	.148**	1	.077	.214**	.300**	.243**	.155**
	Sig. (2-tailed)	.000	.000	.372	.550	.007	.000	.000	.008		.165	.000	.000	.000	.005
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
PA_3	Pearson Correlation	.182**	.277**	.152**	.142*	.177**	.198**	.154**	.099	.077	1	.037	.046	.120*	.065
	Sig. (2-tailed)	.001	.000	.006	.011	.001	.000	.005	.074	.165		.503	.412	.031	.246
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
PA_4	Pearson Correlation	.309**	.308**	.059	.086	.201**	.201**	.380**	.178**	.214**	.037	1	.335**	.244**	.093
	Sig. (2-tailed)	.000	.000	.293	.120	.000	.000	.000	.001	.000	.503		.000	.000	.093
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
PA_5	Pearson Correlation	.286**	.264**	.080	.124*	.241**	.277**	.344**	.205**	.300**	.046	.335**	1	.226**	.149**
	Sig. (2-tailed)	.000	.000	.153	.025	.000	.000	.000	.000	.000	.412	.000		.000	.007
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
PA_6	Pearson Correlation	.331**	.365**	.037	.051	.211**	.349**	.428**	.210**	.243**	.120*	.244**	.226**	1	.292**
	Sig. (2-tailed)	.000	.000	.504	.362	.000	.000	.000	.000	.000	.031	.000	.000		.000
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324
PA_7	Pearson Correlation	.245**	.323**	.084	.132*	.216**	.262**	.297**	.188**	.155**	.065	.093	.149**	.292**	1
	Sig. (2-tailed)	.000	.000	.131	.017	.000	.000	.000	.001	.005	.246	.093	.007	.000	
	N	324	324	324	324	324	324	324	324	324	324	324	324	324	324

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

4.5.1 Form A

All of the Form A subsections showed significant moderate positive correlations with each of the seven PA exercises, providing additional internal validity evidence supporting the single construct hypothesis. PA exercise 1 (Plate Alignment) had the strongest positive correlation with the Problem Solving subsection of the Form A test. PA exercise 2 (Cylinder Alignment) also had the strongest positive correlation with the Problem Solving subsection of the Form A test. PA exercise 3 (Automatic Sequence) had the strongest positive correlation with the Mechanical subsection of the Form A test. PA exercise 4 (Pneumatic System - Vacuum) had the strongest positive correlation with the Pneumatics subsection of the Form A test. PA exercise 5 (Pneumatic System – Cylinder Speed) had the strongest positive correlation with the Electrical subsection of the Form A test. PA exercise 6 (Component Connection) had the strongest positive correlation with the Mechanical subsection of the Form A test. PA exercise 7 (Electrical Circuit Test) had the strongest positive correlation with the Electrical subsection of the Form A test.

4.5.2 Form B

Unlike Form A, not all of the Form B subsections correlated significantly with each of the seven PA exercises. The Hydraulics subsection of Form B did not correlate significantly at the α=.05 level with PA exercise 2 (Cylinder Alignment), PA exercise 4 (Pneumatic System – Vacuum), PA exercise 5 (Pneumatic System – Cylinder Speed), PA exercise 6 (Component Connection), and PA exercise 7 (Electrical Circuit Test). The Hand/Measuring Tools subsection of Form B did not correlate significantly at the α=.05 level with PA exercise 1 (Plate Alignment), PA exercise 2 (Cylinder Alignment), PA exercise 4 (Pneumatic System – Vacuum), and PA exercise 6 (Component Connection).

PA exercise 1 (Plate Alignment) had the strongest positive correlation with the Math/Statistics subsection of the MC Test. PA exercise 2 (Cylinder Alignment) had the strongest positive correlation with the Pneumatics subsection of the Form B test. PA exercise 3 (Automatic Sequence) had the strongest positive correlation with the Electrical subsection of the Form B test. PA exercise 4 (Pneumatic System - Vacuum) had the strongest positive correlation with the Pneumatics subsection of the Form B test, just as it did with the Form A test. PA exercise 5 (Pneumatic System – Cylinder Speed) had the strongest positive correlation with the Pneumatics subsection of the Form B test. PA exercise 6 (Component Connection) had the strongest positive correlation with the Pneumatics subsection of the Form B test. PA exercise 7 (Electrical Circuit Test) had the strongest positive correlation with the Electrical subsection of the Form B test, just as it did with the Form A test.

The correlational results for Form A were compared to the content analysis ratings of the subject-matter experts. Just as the correlations between corresponding MC test subsections and PA exercises tended to have higher, positive correlations, the same was true of the mean relevance ratings from the content experts. Specifically, PA4 (Pneumatic System – Vacuum) had the highest mean relevance rating (1.14) with the Pneumatics subsection of the MC test and they had correlation of .382. PA5 (Pneumatic System – Cylinder Speed) had the highest mean relevance rating (1.36) with the Pneumatics subsection of the MC test and they had a correlation of .350. PA6 (Component Connection) had the highest mean relevance rating (0.76) with the Pneumatics subsection of the MC test and they had a correlation of .228. PA7 (Electrical Circuit Test) had the highest mean relevance rating with the Electrical subsection of the MC test and they had a correlation of .267.

5.0 SUMMARY AND CONCLUSIONS

The purpose of this study was to provide validity evidence for two alternate equivalent multiple-choice (MC) job knowledge tests and the seven performance assessment (PA) exercises that were developed for the L2/L3 Production Technician at a large consumer products manufacturing company in the southeastern United States. Of central interest in this study was whether the PA exercises were measuring the same or additional knowledge, skills, and abilities as the MC tests. Also of primary interest was whether the results of the content analysis, which examined the relationships between the MC and PA tests, were consistent with the results of the other empirical analyses in this study. Each element of this study will be discussed separately.

5.1 CONTENT VALIDITY EVIDENCE FOR THE MC JOB KNOWLEDGE TEST AND THE PA EXERCISES

The purpose of this part of the study was to examine the two subject-matter experts' evaluation of the content relatedness of one form of the MC test and the PA exercises. The subject-matter experts evaluated each MC item in terms of its content and its relationship to each of the seven PA exercises according to a 4-point Likert scale. After each of the MC items was rated by the subject matter experts, the ratings were averaged to determine the mean relevance rating of each of the seven PA exercises to the total MC test. The results were compiled and analyzed by (a) total MC test with total on PA exercises, (b) total MC test with each PA exercise, and (c) each MC subtest with each PA exercise.

The G-coefficient value of 0.32 was lower than expected and may have been a result of several factors. First, the low G-coefficient may be due to the small number of raters that were used in this study. Additionally, both raters viewed many MC item to PA exercise relationships as either “0 = no relationship” or “1 = small relationship” rather than “2 = moderate relationship” or “3 = strong relationship”. This was surprising especially in light of the strong positive correlation between examinee performance on both forms of the MC test and the PA exercises.

As part of the content analysis, the two subject-matter experts were asked to identify for each of the PA exercises if any additional knowledge, skills, or abilities are being assessed beyond what is measured by the MC test items. The responses from the two job experts indicated that while several of the PA exercises are closely related to some of the MC items in terms of content, the PA exercises measure a more applied understanding of that content. This was expected since the MC tests were designed to measure an examinee's knowledge of job specific information, whereas the PA exercises were designed to measure an examinee's ability to perform specific job relevant tasks. Although the MC test and the PA exercises likely measure related constructs, the PA exercises were designed to measure more complex job skill requirements with a distinctly different measurement method than their paper-and-pencil counterpart.

Finally, the subject-matter experts were asked to rate the importance of the seven PA exercises in terms of their importance to the job of L2/L3 Production Technician according to the following scale: (a) 0 = not important (b) 1 = small importance, (c) 2 = moderate importance, and (d) 3 = great importance. Both job experts rated all of the PA exercises a 3 = great importance.

This result was expected since the PA exercises, while reflecting tasks which are somewhat abstracted from actual job duties, were revealed by job and task analysis data, as well as evaluation by job experts, to reflect critical and frequently performed job skill requirements of the L2/L3 Production Technician.

The results of the subject-matter experts' content analysis and ratings were averaged and compiled for each of the MC test sections with each of seven PA exercises. Although most mean relatedness ratings averaged less than 1, it was expected that certain subsections of the MC test would have the highest mean relatedness ratings with their corresponding PA exercises. For example, PA4 (Pneumatic System – Vacuum) had the highest mean relevance rating (1.14) with the Pneumatics subsection of the MC test. Not surprisingly, the PA5 (Pneumatic System – Cylinder Speed) had the highest mean relevance rating (1.36) with the Pneumatics subsection of the MC test. PA6 (Component Connection) had the highest mean relevance rating (0.76) with the Pneumatics subsection of the MC test. As expected, PA7 (Electrical Circuit Test) had the highest mean relevance rating (0.72) with the Electrical subsection of the MC test.

PA1 (Plate Alignment) had the highest mean relevance rating (0.65) with the Mechanical subsection of the MC test. PA2 (Cylinder Alignment) had the highest mean relevance rating with the Mechanical (0.68) and the Hand Tools (0.69) subsections of the MC test. PA3 (Automatic Sequence) had the highest mean relevance rating (0.92) with the Problem Solving subsection of the MC test.

The Mechanical and Pneumatics subsections of the MC test had the highest overall mean relevance ratings (0.58 and 0.60 respectively) with the seven PA exercises. The Math/Statistics subsection of the MC test had the lowest overall mean relevance ratings of 0.16 with the seven PA exercises.

In general, the results of the content analysis supported the hypothesis that subsections of the MC test would have the highest mean relatedness ratings with their correspondingly labeled or titled PA exercise.

5.2 ASSESSING ITEM AND TEST PROPERTIES OF THE MC JOB KNOWLEDGE TESTS AND THE PA EXERCISES

Harris and Crouse (1993) identify four conditions for equating that they attribute to Lord (1980). According to Lord (1980), equity as it applies to the current study, means that it does not matter to each examinee whether they take Form A or Form B. When the two tests are perfectly parallel, the equity property will hold making equating unnecessary.

The results of the item analyses for the two forms of the MC test indicated very similar means, standard deviations, and reliabilities. The means for Form A and Form B were 95.637 and 94.512 respectively. The Standard deviation was 13.856 for Form A and 14.908 for Form B. The coefficient alphas for Form A and Form B were .913 and .921 respectively, indicating excellent reliability. The histograms, as well as the skewness and kurtosis statistics, revealed very similar distributions for Form A and Form B.

The coefficient alpha for the seven PA exercises was .594 which is acceptable considering the small number of items included. Item difficulty, discrimination indices, and point biserial correlations were calculated for each item and exercise. For Form A, seventeen items had point biserial correlations below .20 and Form B, thirteen items had point biserial correlations below .20. The low point-biserial correlations of the non-loading items could be an indication of flawed or ineffective items. Most likely the items with low point-biserial correlations were too easy for this group of examinees as the majority of these items had high proportion/percentage correct statistics.

5.3 INTERNAL STRUCTURE EVIDENCE FOR FORM A AND FORM B WITH THE PA EXERCISES INCLUDED

The results of the Form A CFA analyses revealed that with the exception of the RMSEA fit statistics for Model 1 (which hypothesized that there was one factor underlying the MC and PA test scores) and Model 2 (which hypothesized that there were two factors underlying the MC and PA test scores), the results failed to demonstrate fit of the model to the data. The RMSEA value was identical 0.035 for both Model 1 and Model 2. There was no convergence for Model 3 (which hypothesized that there were eight factors underlying the MC and PA test scores) as the number of iterations was exceeded.

The results of the Form B CFA analyses revealed that with the exception of the RMSEA fit statistic for Model 3, the results for Model 1, Model 2, and Model 3 failed to demonstrate fit of the model to the data.

Because the CFA Results were somewhat ambiguous, further analysis of the factorial structure of the two forms of the MC Test combined with the 7 PA exercises was undertaken by conducting several exploratory factor analyses. For both Form A and Form B, the scree plots and fit statistics supported the appropriateness of a one-factor solution. The single factor solution supports the hypothesis that both forms of the MC test (along with the PA exercises) measure a construct labeled as *applied mechanical knowledge*.

5.4 EXTERNAL VALIDITY EVIDENCE FOR MC JOB KNOWLEDGE TESTS

5.4.1 Relationship Between MC Job Knowledge Tests and Total Score on the PA Exercises

The correlations for MC total score with PA total score were very similar for both Form A and Form B (.627 with .612 respectively). The MC subsections all showed significant positive correlations with PA total score for both test forms. Furthermore, the correlations for each MC subsection with PA total score were of similar magnitude for both forms with the exception of Hydraulics (.484 versus .153 for Form A and Form B respectively) and Hand/Measuring Tools (.323 versus .162 for Form A and Form B respectively). The linear regression results for Form A and Form B were very similar for both forms and there were no apparent assumption violations.

Decision consistency of both forms of the MC test with the PA exercises was also examined and compared. Decisions for an examinee are consistent when the results of both the MC test and the PA test indicate that an examinee should be classified as passing. The percentage of consistent classifications from the MC test and the PA test was calculated by summing the percentage of examinees that passed both the MC test and the PA test with the percentage of examinees who failed both tests. The decision consistency for Form A was 0.70 and the decision consistency for Form B was 0.73. The probability of an inconsistent decision was 0.30 for Form B and 0.27 for Form B. Thus, both forms of the MC test resulted in similarly consistent decisions with the seven PA exercises. The decision consistency is probably not high enough to support the use of either the MC test or the PA exercises alone to select L1/L2 Production Technicians.

5.4.2 Relationship Between Each Subtest of the MC Job Knowledge Test and Each of the Seven PA Exercises

All of the Form A subsections showed significant moderate positive correlations with each of the seven PA exercises. However, unlike Form A, not all of the Form B subsections correlated significantly with each of the seven PA exercises. The smaller sample size for Form B may have contributed to the nonsignificant correlations.

The correlational results for Form A were compared to the content analysis ratings of the subject-matter experts. As expected, the correlations between correspondingly labeled or titled MC test subsections and PA exercises tended to have higher, positive correlations. The mean relevance ratings from the content experts also tended to be higher where it was anticipated (e.g., PA7-Electrical Circuit Test had the highest mean relevance rating with the Electrical subsection of the MC test).

5.5 CONCLUDING REMARKS

The purpose of this study was to examine the validity evidence for two alternate multiple-choice (MC) job knowledge tests and seven performance assessment (PA) exercises that were developed for employment selection purposes. Of central importance to this study was whether the PA exercises were providing substantial additional information beyond the MC tests regarding the examinees' knowledge, skills, and abilities. Because of the additional administration time and financial costs associated with the use of the PA exercises, if the MC tests were found to measure the same construct(s) with same effectiveness of the PA exercises, then using only the MC test without the PA exercises could be justified. To the extent that they can be generalized, the results of this study have implications not only for the L2/L3 Production

Technician but also for other employers who must consider the additional expenses associated with the development, validation and administration of performance tests as part of a selection procedure. While PA measures can often assess more complex job skill requirements than paper-and-pencil MC tests, the value and the amount of additional information gained from PA measures must be weighed against the costs.

The evidence collected in this study appears to support the idea that the same construct labeled as *applied mechanical knowledge* is being measured by both the MC tests and the PA exercises. Additionally the evidence supports the use of both Form A and Form B as alternate test forms. However, the decision consistency between the MC tests and the PA exercises does not appear to be sufficient to recommend that either form of the MC test alone could be used to select qualified L2/L3 Production Technicians. It is likely that a considerable amount of information regarding an examinee's ability is "lost" or not captured due to the fact that the PA exercises use a dichotomous scoring rubric. In the future, serious consideration should be given to using a polytomously scored format for performance tests in employment settings.

While several key hypotheses of this study were supported, some of the content analysis results were unexpected. It was expected that results of the content analysis, which examined relationships between the MC and PA tests, would be consistent with results from the other empirical analyses in this study. For example, it was predicted that the content analysis would show an overall strong relationship between the MC items and the PA exercises. However, this was not the case as the subject-matter experts rated a much larger than expected number of MC items as either having "no relationship" or "small relationship" to the PA exercises. Perhaps when examined at the individual item level, the relationship between a particular MC item and a particular PA exercise was not readily apparent. It may be that the MC test must be considered in its entirety or perhaps by test subsections in order to perceive the content relationships with the PA exercises. In fact, one subject-matter expert's comments following his content analysis support this idea.

In spite of the content analysis data, overall evidence showed a link between MC items and PA exercises. Moreover, the subject-matter experts found a very strong link between the PA exercises and the job of L2/L3 Production Technician as they both rated each PA exercise as having "great importance" to the job. However, additional research may be necessary to examine the actual job performance of those examinees that passed both the MC items and the PA exercises as part of the original selection procedure. If the PA exercises are no longer included as part of the selection procedure, additional research could compare the job performance of those who took both selection instruments with those who passed only the MC test.

APPENDIX A

KNOWLEDGE SKILLS AND ABILITIES

L1 (TEAM MEMBER)
KNOWLEDGE, SKILLS & ABILITIES

A. Cognitive Knowledge, Skills and Abilities

*1. Ability to monitor, operate, and adjust first-generation machines and equipment to manufacture, process, and package contact lenses.

*2. Ability to inspect visually products and equipment for large and small defects or errors.

3. Ability to read simple and detailed information in English in procedures, manuals, screens, and communications.

*4. Ability to read and record information from dials and gauges.

*5. Ability to write, enter, and verify figures and information in data sheets or specifications.

*6. Ability to enter data by keyboard into computer system.

7. Ability to manipulate or control objects through hand and arm movements.

8. Ability to perform arithmetic operations including addition, subtraction, multiplication, and division of whole numbers. Ability to read graphs and understand decimals (e.g., ranges).

9. Ability to read a rule or scale to tenths.

10. Ability to follow spoken and written instructions.

11. Ability to communicate orally in English (cordial and professional).

12. Ability to respond and resolve problems quickly.

13. Ability to be alert and observant.

14. Ability to learn and willingness to follow company policies and procedures.

*15. Ability to train others formally and informally.

B. Physical and Related Skills and Abilities

16. Ability to use eye-hand coordination.

17. Ability to walk, stand, sit, lift, bend, reach, push, and climb stairs.

18. Ability to make fine hand and finger movements.

19. Ability to lift magazine with trays (25 lbs.) or canister (2 people).

20. Ability to maintain personal hygiene.

C. Personality, Motivational and Related Abilities

21. Ability to work as scheduled, including shifts.

22. Ability to work as a cooperative team member.

23. Ability to work safely.

24. Ability to demonstrate conscientiousness and a good work ethic.

25. Ability to be flexible and adapt to change.

* After review by L1 job experts on 11-8-99, these knowledge, skills & abilities were determined to be learned on the job.

L2 (PRODUCTION TECHNICIAN)
KNOWLEDGE, SKILLS & ABILITIES

A. Cognitive Knowledge, Skills and Abilities

*1. Ability to setup, operate, monitor, adjust and maintain second-generation lens manufacturing machines and equipment.

2. Ability to use hand tools such as screwdrivers, wrenches, and sockets.

3. Ability to read simple and detailed manuals, procedures, and screens.

4. Ability to visually inspect products and equipment for errors or defects.

5. Ability to assemble and disassemble machines and equipment (electrical & pneumatic).

6. Ability to use a working familiarity with electromechanical technology to setup, operate, maintain, and repair a complex second-generation lens fabrication system.

7. Ability to lubricate machines and equipment.

8. Ability to clean up machines and work area.

9. Ability to read and record information from counters, charts, graphs, gauges, dials, and screens.

10. Ability to perform various numerical operations including addition, subtraction, multiplication, and division of whole numbers and decimals; percentages; means; positive and negative numbers; and minimum and maximums.

11. Ability to receive simple and complex information from conversation and instructions.

12. Ability to write figures and detailed information in data sheets, reports, or logs.

13. Ability to solve simple and complex problems including troubleshooting and diagnosis.

14. Ability to use judgment beyond written or oral instructions.

15. Ability to combine information from several sources to make decisions.

16. Ability to break down information or data into component parts, such as analyzing production problems.

17. Ability to communicate orally and in writing in English.

18. Ability to use PC (email, Word, plant information system) and keyboard to operate equipment.

19. Ability to measure accurately to 1/10 millimeter or 50 microns.

*20. Knowledge of GMP and FDA requirements.

- Continued -

Continued

L2 (PRODUCTION TECHNICIAN)
KNOWLEDGE, SKILLS & ABILITIES

21. Knowledge of company policies and procedures.

22. Ability to read assembly drawings.

B. Physical and Related Skills and Abilities

23. Ability to walk, climb, stoop, bend, reach, and lift 25 lbs. (pallets, cartons, equipment, product).

24. Ability to make fine hand and finger movements.

C. Personality, Motivational and Related Abilities

25. Ability to work in a team environment.

*26. Ability to work safely around hazardous chemicals and equipment.

27. Flexibility and ability to adapt to change.

28. Ability to work as scheduled and overtime (12 hour shift).

29. Ability to work with minimal supervision.

* After review by L2 job experts on 11-8-99, these knowledge, skills & abilities were determined to be learned on the job.

L3 (Production Technician)
KNOWLEDGE, SKILLS & ABILITIES

A. Cognitive Knowledge, Skills and Abilities

*1. Ability to set up, operate, monitor, evaluate, adjust, and maintain machines and equipment in a 3rd generation contact lens production facility.

2. Ability to operate computer or other systems for input/output, email, boot, and manipulate files.

3. Ability to use hand tools such as screwdrivers, wrenches and sockets.

4. Ability to read simple and detailed instructions or information in manuals, procedures, and screens.

5. Ability to inspect visually for errors or defects in products and materials.

6. Ability to assemble and disassemble machines and equipment.

7. Ability to use a working familiarity with a body of electrical and mechanical knowledge at the technology level.

8. Ability to lubricate and clean up machines and equipment.

9. Ability to read or report and record information from counters, charts, graphs, gauges, dials, or other similar devices.

10. Ability to give and receive simple and complex job information and instruction orally in English.

11. Ability to write, copy material, fill out logs or data sheets, and compose simple written communications in English (logs or work reports).

12. Ability to solve problems using general information as in troubleshooting production problems.

13. Ability to exercise judgment and initiative beyond oral instructions.

14. Ability to perform arithmetic operations including addition, subtraction, multiplication, and division of whole numbers, fractions and decimals; make conversions (English, metric, bars, PSI), and percentages.

15. Ability to measure accurately to ten-thousandths and 10 microns.

16. Knowledge of electrical, hydraulic, and pneumatic systems.

17. Ability to read prints, schematics, flow diagrams, and basic understanding of structural program language.

18. Ability to organize information and attend to detail.

** *19. Ability to learn and apply a knowledge of process technology.

B. Physical and Related Skills and Abilities

20. Ability to walk, stand, bend, stoop, crawl, climb, reach, and lift (25 lb. foil).

21. Ability to handle, insert, and turn screws (finger & manual dexterity).

- Continued -

Continued

L3 (Production Technician)
KNOWLEDGE, SKILLS & ABILITIES

22. Ability to perform routine and repetitive tasks.
23. Ability to work in yellow lighting.

C. Personality, Motivational and Related Abilities

24. Ability to work as a cooperative member of a team.
25. Ability to work safely and efficiently.
26. Ability to work as scheduled (12 hour shift) including rotating shifts and overtime.
27. Conscientious and dedicated in job performance.

* After review by L3 job experts on 11-8-99, these knowledge, skills & abilities were determined to be learned on the job.

** Added by L3 job experts on 11-8-99.

APPENDIX B

JOB ACTIVITY CHECKLIST

Marking Instructions

Job Activity Checklist

- Use a No. 2 pencil only.
- Do not use ink, ballpoint, or felt tip pens.
- Make solid marks that fill the oval completely.
- Erase cleanly any marks you wish to change.
- Make no stray marks on this form.
- Do not fold, tear, or mutilate this form.

INCORRECT ✓ ✗ ⊖ ⊙ CORRECT ●

Name of Rater | Date of Rating

Company Name and Location

Rater's Job Title

Department and Phone Number

Title of Job Rated

Name of Training You Completed
Degree (specify)
Assoc: ○ ______ Bacc: ○ ______ Master's: ○ ______
OJT: ○ ______ Apprenticeship: ○ ______ Other: ○ ______

Years of Incumbency or Supervision For Job You Are Rating I: S:

Rater's Signature

Directions

Each item in this list refers to something a person might do on a job. For each item you are to rate the importance of the item to the job listed on this page. In rating importance you should consider how much the activity applies to the job and how critical the activity is to the overall performance of the job. Consider the time on a weekly basis. Carefully examine each item and then mark one of the three ovals as shown below:

Importance — EXAMPLE X

Does not apply / Important

● ① ② Does not apply. The 0 is marked because the activity is not done at all by a person on this job.

⓪ ● ② Done by a person as part of the job. The 1 is marked because the activity is done by a person on this job, but it is not one of the more important parts of the job.

⓪ ① ● Important. The 2 is marked because the activity is an important part of the job and is necessary to perform the job.

Many of the items contain examples. These examples are there only to make the item clearer. Don't be limited by the examples in deciding whether or not something is done on the job you are considering. Some jobs may involve many of the activities while others may involve fewer. Then, for each cluster of duties, mark the percent of time spent on those duties.

EXAMPLE Y

A. Operates Stationary Machines → 0 5 Percent of Time.

The 5 indicates this person spends 5 percent of his or her time operating stationary machines. When you have done this for each cluster, the total percents of time must add up to 100 percent.

Remember: Consider only the job being analyzed. You are to look at the job, not the person.

Published by
ROLAND T. RAMSAY, Ph.D.
RAMSAY CORPORATION
1050 BOYCE ROAD
PITTSBURGH, PA 15241-3907

A. Operates Machines or Equipment → Percent of Time.

Importance

Does not apply / Important

⓪ ① ② 1. Operates or uses stationary machines or equipment to process, fabricate, or otherwise change parts, objects, or materials.

⓪ ① ② 2. Monitors and adjusts controls for an operation or process.

⓪ ① ② 3. Sets up or adjusts machines or equipment.

⓪ ① ② 4. Operates computer or other system.

⓪ ① ② 5. Operates nonpowered mobile equipment; such as hand truck, dolly, or wheelbarrow.

⓪ ① ② 6. Operates equipment in motion; such as crane or conveyor system.

⓪ ① ② 7. Operates powered land vehicles intended for highway or railroad transportation; such as automobile, truck, or bus.

⓪ ① ② 8. Operates powered mobile equipment typically not intended for highway or railroad use; such as scooter, warehouse truck, or forklift.

PLEASE DO NOT WRITE IN THIS AREA

959

B. Uses Hand and Powered Tools Associated with the Crafts — Percent of Time. (0–9, 0–9)

Importance — Does not apply / Important

0	1	2	
⓪	①	②	9. Uses hand tools operated by hand power to perform accurate operations; such as fitting or polishing (do not include use of measuring instruments as in items 25 and 26).
⓪	①	②	10. Uses nonprecision tools to perform operations not requiring great accuracy; such as hammer, wrench, trowel, or putty knife.
⓪	①	②	11. Uses hand tools for making adjustments; such as screwdriver, wrench, or file.
⓪	①	②	12. Uses long-handled tools to perform operations not requiring great accuracy or precision; such as pinch bar, shovel, broom, or mop.
⓪	①	②	13. Uses instruments or devices for sketching, illustrating, or drafting; such as pen, pencil, drawing instrument, or CAD system.
⓪	①	②	14. Uses powered, hand-held tools or implements to perform operations requiring accuracy; such as pencil grinder, drill, or miniature soldering iron.
⓪	①	②	15. Uses powered hand tools or implements to perform operations not requiring great accuracy or precision; such as sander or chipping hammer.

C. Reads Instructions, Texts, Manuals, or Reports — Percent of Time. (0–9, 0–9)

Importance — Does not apply / Important

0	1	2	
⓪	①	②	16. Reads simple instructions or information; such as safety signs or work schedules.
⓪	①	②	17. Reads detailed instructions or information; such as textbooks, work procedures, or equipment manuals.
⓪	①	②	18. Reads detailed complex instructions or information; such as complex equipment manuals or assembly descriptions.

D. Inspects, Locates Problems, and Corrects Defects or Breakdowns — Percent of Time. (0–9, 0–9)

Importance — Does not apply / Important

0	1	2	
⓪	①	②	19. Inspects objects or materials to detect small errors or defects; such as inspection of damaged threads or locating defects in highly polished surfaces.
⓪	①	②	20. Inspects objects to detect large errors or defects; such as major imperfections in materials.
⓪	①	②	21. Assembles or disassembles objects; such as puts parts or replacements together to build or form an object, or takes apart components.
⓪	①	②	22. Uses a working familiarity with a body of technical knowledge; such as chemical, electronic, electrical, optical, mechanical, metallurgical, or other technology. Please specify one or more by writing in the technology: ____________

E. Reads Blueprints, Schematics, Diagrams, or Charts — Percent of Time. (0–9, 0–9)

Importance — Does not apply / Important

0	1	2	
⓪	①	②	23. Reads information coded alphabetically, numerically, or symbolically.
⓪	①	②	24. Reads sketches, simple blueprints, drawings, diagrams, or schematics.
⓪	①	②	25. Reads and interprets complex blueprints or drawings; such as detailed assembly and parts drawings.

F. Performs Activities to Prevent Equipment Wear or Failure — Percent of Time. (0–9, 0–9)

Importance — Does not apply / Important

0	1	2	
⓪	①	②	26. Lubricates machines or equipment.
⓪	①	②	27. Cleans up machines and work area.

Mark Reflex® by NCS EM-213833-1:65432 Printed in U.S.A.

G. Measures with Scale, Tape, or Other Instruments

Percent of Time.

Importance (Does not apply — Important)

⓪ ① ② 28. Measures with simple measuring devices; such as tape or scale.

⓪ ① ② 29. Measures with complex or close tolerance devices; such as calipers, micrometer, etc.

⓪ ① ② 30. Reads and reports or records information from counters, charts, graphs, cumulative recorder tracings, calibrated gauges, dials, or other similar devices.

⓪ ① ② 31. Measures or takes readings with complex mechanical, electrical, chemical, or optical devices; such as pyrometer, optical comparator, oscilloscope, laser system, or universal testing machine.

H. Performs Arithmetic Operations

Percent of Time.

Importance (Does not apply — Important)

⓪ ① ② 32. Arranges material in a given order by alphabetizing, matching, or placing in numerical sequence.

⓪ ① ② 33. Adds or subtracts simple whole numbers of two or fewer digits.

⓪ ① ② 34. Multiplies and divides whole numbers or adds and subtracts whole numbers of three or more digits.

⓪ ① ② 35. Adds, subtracts, multiplies, or divides decimals or percentages.

⓪ ① ② 36. Adds and subtracts fractions.

⓪ ① ② 37. Multiplies and divides fractions.

⓪ ① ② 38. Uses tables; such as converting from size to weight, or from fractions to decimals or vice versa.

⓪ ① ② 39. Uses simple formulas; such as converting from size to weight or vice versa.

⓪ ① ② 40. Calculates conversion from fractions to decimals.

⓪ ① ② 41. Obtains square roots or cube roots with or without calculator.

⓪ ① ② 42. Calculates areas, volumes, or quantities.

⓪ ① ② 43. Uses geometric principles in solving problems.

⓪ ① ② 44. Solves problems using trigonometric functions.

I. Receives Instruction

Percent of Time.

Importance (Does not apply — Important)

⓪ ① ② 45. Receives simple information from job-related oral communication; such as oral instructions, conversations, orders, requests, interviews, and meetings.

⓪ ① ② 46. Receives complex or technical information orally, in conversation or in meetings; such as a discussion of design problems or receiving informal instruction on technical matters.

J. Communicates in Writing, Orally, or on Keyboard

Percent of Time.

Importance (Does not apply — Important)

⓪ ① ② 47. Writes name or copies written material.

⓪ ① ② 48. Writes figures or simple information according to written or oral instructions or from own observation; such as filling out data sheets.

⓪ ① ② 49. Codes information using numerical or alphabetical codes (except color codes), drafting symbols, replacement part numbers, or catalog numbers.

⓪ ① ② 50. Writes detailed information as in work performed or observer reports.

⓪ ① ② 51. Enters data by keyboard into computer or video display unit.

⓪ ① ② 52. Composes simple written communications; such as work reports or logs.

⓪ ① ② 53. Composes detailed complex written communications; such as research reports or summary reports.

⓪ ① ② 54. Makes oral reports of developments or results of job in progress.

K. Estimates, Solves Problems, or Makes Decisions

Importance (Does not apply ⓪ ① ② Important) — Percent of Time.

- ⓪ ① ② **55.** Estimates weight of objects.
- ⓪ ① ② **56.** Estimates size, length, or quantities involving small objects; such as miniature parts.
- ⓪ ① ② **57.** Estimates size, length, or quantities involving large objects; such as wrench for bolts or size of fittings.
- ⓪ ① ② **58.** Estimates sizes of objects or areas relative to sizes of other objects or areas; such as parking a vehicle or placing a lift in a designated area.
- ⓪ ① ② **59.** Makes low-level decisions; such as storing parts or sorting objects.
- ⓪ ① ② **60.** Makes higher-level decisions which influence the work of others; such as deciding the sequence of events to take place.
- ⓪ ① ② **61.** Solves problems by selecting the correct solution from a small number of options; such as selecting proper materials or selecting the correct tool for a given task.
- ⓪ ① ② **62.** Solves problems using general information; such as trouble-shooting production problems.
- ⓪ ① ② **63.** Exercises considerable judgment and initiative beyond written or oral instructions.
- ⓪ ① ② **64.** Evaluates a continuing process and adjusts machine or instruments to it.
- ⓪ ① ② **65.** Combines information from two or more sources to establish new facts or basis for decision; such as evaluating the effects of temperature stress on a part.
- ⓪ ① ② **66.** Breaks down information or data into component parts; such as analyzing production problems.
- ⓪ ① ② **67.** Gives simple work direction or instruction to other workers on either a formal or an informal basis; such as directing or instructing a novice.

L. Uses Hand or Fingers in Accomplishing Tasks

Importance (Does not apply ⓪ ① ② Important) — Percent of Time.

- ⓪ ① ② **68.** Arranges or positions objects or materials in an orderly arrangement or specific position; such as stocking shelves, filing materials, warehousing or storing materials.
- ⓪ ① ② **69.** Handles objects or materials by hand or with simple helping instruments.
- ⓪ ① ② **70.** Manipulates or controls objects through hand or arm movements; such as placing equipment or packaging materials.
- ⓪ ① ② **71.** Maintains uniform, controlled, or steady hand-arm posture or movement; such as using a welding or burning torch or pouring liquid into a small vessel.
- ⓪ ① ② **72.** Makes careful finger movements in various types of activities; such as assembly of small parts or use of precision tools.
- ⓪ ① ② **73.** Uses hands directly to form or shape material or products; such as making a mold or shaping warm metal.
- ⓪ ① ② **74.** Moves hands or feet in coordination with what the eyes see; such as driving a vehicle or operating a shear.
- ⓪ ① ② **75.** Manually controls or guides materials being processed; such as guiding material in a machine.
- ⓪ ① ② **76.** Manually inserts material in or removes from machines or equipment; such as operation of punch press.

M. Miscellaneous

Importance (Does not apply ⓪ ① ② Important) — Percent of Time.

- ⓪ ① ② **77.** Works near hazardous materials and surroundings.
- ⓪ ① ② **78.** Works cooperatively as a member of a team.
- ⓪ ① ② **79.** ____________________
- ⓪ ① ② **80.** ____________________

PLEASE DO NOT WRITE IN THIS AREA

959

APPENDIX C

SUMMARY OF JOB ACTIVITY CHECKLIST RESULTS

The Job Activity Checklist (JAC) was designed by Ramsay (1970) to suggest criteria for development of selection procedures in manufacturing and processing. A copy of the JAC is included in Appendix B. Job experts (raters) were asked to indicate whether a given task was: 2, important; 1, done but not one of the most important parts of the job; or 0, not done by a person on the job. In addition, each rater estimated the percent of time an incumbent would spend on various tasks.

In order to determine what constituted a significant number of job experts, Lawshe's (1975) Content Validity Ratio was calculated. Significance was evaluated using Wood's (undated) Table 2. For attainment of the .05 level of significance a minimum of 8 of 9, 9 of 11, or 7 of 8 raters must agree that an item is essential. Table 29 shows the numbers of JAC items that were significant at the .05 level for the three jobs.

Table 29. JAC Items Significant at the .05 Level

L1 (Team Member)	L2 (Production Technician)			L3 (Production Technician)		
1	1	26	52	1	22	52
2	2	27	61	2	27	62
16	3	30	62	4	30	63
17	11	34	63	16	45	64
30	16	35	64	19	47	78
48	17	45	65	21	48	
51	19	46	66			
64	20	48	78			
70	21	50				
78	22	51				

For the JAC the intraclass correlation of mean ratings was computed as described by Guilford & Fruchter (1978). Table 30 below shows the intraclass correlation for L1 (Team Member), L2 (Production Technician) and L3 (Production Technician).

Table 30. Intraclass Correlations for the Three Jobs

Job	Number of Raters	Intraclass Correlation
L1 (Team Member)	9	.92
L2 (Production Technician)	11	.93
L3 (Production Technician)	8	.88

The data in Table 32 show that the raters were in agreement on the tasks performed by the three jobs.

APPENDIX D

SUBJECT-MATTER EXPERT SURVEY WITH INSTRUCTIONS

Survey Part 1

For this part of the survey, please review each item (starting with item 3) on the *L2/L3 Production Technician Test – Form A* and rate the degree of relationship between the item's content and the content of each of the seven PA exercises according to the key in the upper-right-hand corner of the following page.

Rater Code A2

Date Rated

Key
3 = strong relationship
2 = moderate relationship
1= small relationship
0 = no relationship

MC Item #	PA 1 Plate Alignment	PA 2 Cylinder Alignment	PA 3 Automatic Sequence	PA 4 Pneumatic System (Vacuum)	PA 5 Pneumatic System (Cylinder Speed)	PA 6 Component Connection	PA 7 Electrical Circuit Test
3							
4							
5							
6							
7							
8							
9							
10							
11							
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31							
32							
33							
34							
35							
36							
37							
38							
39							
40							
41							
42							
43							
44							

MC Item #	PA 1 Plate Alignment	PA 2 Cylinder Alignment	PA 3 Automatic Sequence	PA 4 Pneumatic System (Vacuum)	PA 5 Pneumatic System (Cylinder Speed)	PA 6 Component Connection	PA 7 Electrical Circuit Test
45							
46							
47							
48							
49							
50							
51							
52							
53							
54							
55							
56							
57							
58							
59							
60							
61							
62							
63							
64							
65							
66							
67							
68							
69							
70							
71							
72							
73							
74							
75							
76							
77							
78							
79							
80							
81							
82							
83							
84							
85							
86							

MC Item #	PA 1 Plate Alignment	PA 2 Cylinder Alignment	PA 3 Automatic Sequence	PA 4 Pneumatic System (Vacuum)	PA 5 Pneumatic System (Cylinder Speed)	PA 6 Component Connection	PA 7 Electrical Circuit Test
87							
88							
89							
90							
91							
92							
93							
94							
95							
96							
97							
98							
99							
100							
101							
102							
103							
104							
105							
106							
107							
108							
109							
110							
111							
112							
113							
114							
115							
116							
117							
118							
119							
120							
121							
122							
123							
124							
125							
126							
127							

Survey Part 2

For this part of the survey, please review each of the seven PA exercises and rate their importance to the job of L2/L3 Production Technician according to the scale on the following page. Additionally, on the following page, please list or describe any additional Knowledge, Skills, and Abilities that you feel are being assessed by the PA exercises that are beyond those that are being measured by the multiple-choice test items.

Rater Code ____________________
Date Rated ____________________

Name	Importance to the Job (Circle One)	Please list or describe any additional Knowledge, Skills, and Abilities that are being assessed by these exercises that are beyond those that are being measured by the multiple-choice test items.
1. Plate Alignment	**3** = great importance **2** = moderate importance **1** = small importance **0** = not important	
2. Cylinder Alignment	**3** = great importance **2** = moderate importance **1** = small importance **0** = not important	
3. Automatic Sequence	**3** = great importance **2** = moderate importance **1** = small importance **0** = not important	
4. Pneumatic System (Vacuum)	**3** = great importance **2** = moderate importance **1** = small importance **0** = not important	
5. Pneumatic System (Cylinder Speed)	**3** = great importance **2** = moderate importance **1** = small importance **0** = not important	
6. Component Connection	**3** = great importance **2** = moderate importance **1** = small importance **0** = not important	
7. Electrical Circuit Test	**3** = great importance **2** = moderate importance **1** = small importance **0** = not important	

BIBLIOGRAPHY

Aiken, L.R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement, 40*, 955-959.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, D.C: Author.

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

Breland, H.M., & Griswold, P.A., (1982). Use of a performance test as a criterion in a differential validity study. *Journal of Educational Psychology, 74*(5), 713-721.

Brennan, R.L. (1994). Variance components in generalizability theory. In: C.R. Reynolds (Ed), *Cognitive assessment: A multidisciplinary perspective (pp. 175-207).* New York: Plenum Press.

Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing Structural Equation Models*. Newbury Park, CA: Sage Publications.

Callinan, M. & Robertson, I.T. (2000). Work sample testing. *International Journal of Selection and Assessment, 8*(4), 248-260.

Carey, N.B. (1991). Setting standards and diagnosing training needs with surrogate job performance measures. *Military Psychology, 3*, 135-150.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Christensen, L.B., & Stoup, C.M. (1986). *Introduction to statistics for the social and behavioral sciences*. Belmont, CA: Brooks/Cole Publishing Company.

Comrey, A.L., & Lee H.B. (1992). *A first course in factor analysis* (*2nd ed.*). Hillsdale, NJ: Lawrence Erlbaum Associates.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth, TX: Harcourt Brace Javonovich College Publishers.

Dolmans, D.H.J.M., Gijselaers, W.H., & Schmidt, H.G. (1992, April 20-24). Assessing Test Validity Through the Use of Teachers' Judgments. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Downing, S.M. (2003). Validity: on the meaningful interpretation of assessment data. *Medical Education, 37*, 830-837.

Downs, S., Farr, R.M. & Colbeck, L. (1978). Self appraisal: A convergence of selection and appraisal. *Journal of Occupational Psychology, 51*, 271-8.

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

Guilford, J. P., & Fruchter, B. (1978). *Fundamental statistics in psychology and education* (6th ed.). New York: McGraw-Hill.

Hambleton, R.K. (1984). Validating the test score. In R.A. Berk (Ed.), *A guide to criterion-referenced test construction* (199-230). Baltimore: Johns Hopkins University Press.

Hambleton, R.K. & Rogers, J.H. (1988). *Design of an item bias review form: Issues and questions*. Albany, NY: New York State Education Department.

Haladyna, T.M. (1994). *Developing and validating multiple-choice test items*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Harris, D.J., & Crouse J.D. (1993). A study of criteria used in equating. *Applied Measurement in Education, 6*(3):195-240.

Hattrup, K., & Schmitt, N. (1990). Prediction of trades apprentices' performance on job sample criteria. *Personnel Psychology, 43*, 453-466.

Hemphill, J., & Westie, C.M. (1950). The measure of group dimensions. *Journal of Psychology, 29*, 325-342.

Hogan, J., Arneson, S., & Petersons, A.V. (1992). Validation of physical ability tests for high pressure cleaning occupations. *Journal of Business and Psychology, 7*(2), 119-135.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

Joreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika, 34*, 183-202.

Joreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183-202.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141-151.

Kane, M. (2002). Inferences about variance components and reliability-generalizability coefficients in the absence of random sampling. *Journal of Educational Measurement, 39*(2), 165-181.

Kolen, M.J., & Brennan, R.L. (1995). *Test equating: Methods and practices*. New York: Springer-Verlag.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*, 563-575.

Livingston, S.A., & Zieky, M. (1982). *Passing Scores: A manual for setting standards of performance on educational and occupational tests*. Educational Testing Service.

Loehlin, J.C. (1998). *Latent variable models: An introduction to Latent variable models: An introduction to factor, path, and structural analysis* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Long, J.S. (1983). *Confirmatory factor analysis*. Newbury Park, CA: Sage Publications

Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley, 1968.

Lu, K.H. (1971). A measure of agreement among subjective judgments. *Educational and Psychological Measurement, 31*, 75-84.

Messick, S. (1993). Validity. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 13-103). Phoenix: American Council on Education/Macmillan Publishing.

Messick, S. (1996). Validity of performance assessment. In G. W. Phillips (Ed.), *Technical issues in large-scale performance assessment* (pp. 1-18). Washington DC: National Center for Education Statistics.

Millman, J., & Greene, J. (1993). The specification and development of tests of achievement and ability. In R.L. Linn (Ed.), *Educational measurement* (3rd ed.) (pp. 335-366). Phoenix: American Council on Education/Macmillan Publishing.

Muraki, E., Hombo, C.M., & Lee, Y.W. (2000). Equating and linking of performance assessments. *Applied Psychological Measurement, 24* (4), 325-337.

Muthen, L.K., & Muthen, B.O. (1998). *Mplus User's Guide*. Los Angeles: Muthen & Muthen.

Neter, J., Kutner, M.H., Nachtsheim, C.J., & Wasserman, W. (1996). *Applied linear regression models* (3rd ed.). Chicago: Irwin.

Nitko, A.J. (1996). *Educational Assessment of students* (2nd ed.). New York: Macmillan.

Ramsay, R. T. (1970). *Criteria development for test selection through factor analysis of job activity checklist data*. Unpublished doctoral dissertation. University of Pittsburgh.

Ramsay, R.T. (1999). *Content validation report: [L1 (team member)]*. Pittsburgh, PA: Author.

Ramsay, R.T. (2000a). *Content validation report: [L2 (production technician)]*. Pittsburgh, PA: Author.

Ramsay, R.T. (2000b). *Content validation report: [L3 (production technician)]*. Pittsburgh, PA: Author.

Ramsay, R.T. (2000c). *Content validation report: Assessment selection & development [L2&L3 (production technician)]*. Pittsburgh, PA: Author.

Ramsay, R.T. (2003). *The testing manual: A guide to test administration and use.* Pittsburgh, PA: Author.

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

Rovinelli, R.J., & Hambleton, R.K. (1976, April 19-23). *On the Use of Content Specialists in the Assessment of Criterion-Referenced Test Item Validity*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262-274.

Schmidt, N., Clause, C.S., & Pulakos, E.D. (1996). Subgroup differences associated with different measures of some common job-relevant constructs. In C.L. Cooper and I.T. Robertson (eds.), *International Review of Industrial and Organizational Psychology*, Vol. 11. Chichester: Wiley.

Seberhagen, L. W. (1999a). *Position Analysis Questionnaire job analysis for [team member (L1*)]. Vienna, VA: Author.

Seberhagen, L. W. (1999b). *Position Analysis Questionnaire job analysis for [production technician (L2)]*. Vienna, VA: Author.

Seberhagen, L. W. (1999c). *Position Analysis Questionnaire job analysis for [production technician (L3)]*. Vienna, VA: Author.

Shavelson, R.J., & Webb, N.M. (1991). *Generalizability theory: A primer*. Newbury Park, NJ: Sage Publications.

Sireci, S. G. (1995, April). *The central role of content representation in test validity.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Sireci, S.G., & Geisinger, K.F. (1995). Using subject-matter experts to assess content representation: An MDS analysis. *Applied Psychological Measurement, 19(*3), 241-255.

Steiner, D.S., & Gulliland, S.W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology, 81,* 134-141.

Velicer, W. F., & Jackson, D.N. (1990). Component analysis versus common factor-analysis – some further observations. *Multivariate Behavioral Research, 25*(1), 97-114.

Wainer, H. (1989). The future of item analysis. *Journal of Educational Measurement, 26,* 191-208.

Webb, N.M., Rowley ,G.L., & Shavelson, R.J. (1988). Methods, plainly speaking: Using generalizability theory on counseling and development. *Measurement and Evaluation in Counseling and Development, 21,* 81-90.

Wood, R. J.(undated). *Content validation: Procedures and instructions for the determination of item Content Validity Ratios (CVRs) and of the Content Validity Index (CVI) for a test*. East Chicago, IN: Inland Steel Company.

Yalow, E.S., & Popham, W.J. (1983). Content validity at the crossroads. *Educational Researcher, 12*, 10-14.

Yu, C.Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Doctoral dissertation, University of California, Los Angeles.