PXT SELECT™
A Wiley Brand

Customer Service Profile™ Research Report

Version 1.0

WILEY

This edition published in 2020 by John Wiley & Sons, Inc., 400 Highway 169 South, Suite 300, Minneapolis, MN 55426, USA.

Global Headquarters
111 River Street, Hoboken, NJ 07030, USA
For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Printed in the United States of America 10 9 8 7 6 5 4 3 2

Contents

Chapter 1: Introduction

Serving the customer is an important endeavor for all employees regardless of their individual responsibilities. At some point in their career, an employee will likely be in a position to represent his or her company to a customer, client, or other outside agent. While the opportunity to interact with such people may vary from job to job, any outside contact should be a comfortable and satisfying experience for the customer. The impression made by a company's employees often lasts longer than the professional relationship. The impressions made by employees may be characterized by behaviors common to each of us. Assessments designed to identify these behaviors, especially with regard to customer contact, can be helpful for enhancing the quality of any workforce. Those same behaviors are also crucial in developing positive relations between employees within a company.

The Customer Service Profile II™ (CSP II) was developed to help select individuals, regardless of their job titles, who will be successful in fulfilling the service needs of their customers. Various samples of customer service professionals have been studied to identify the most relevant service-oriented traits involved in the world of work today. These results were applied to a greater, more diverse job population representing all types of employees and activities in their organizations.

The CSP II report covers the respondent's Company Service Perspective (one's alignment with the service

orientation of an organization), two basic work-related Proficiencies, and six important Behavioral Traits: Trust, Tact, Empathy, Conformity, Focus, and Flexibility. Additionally, a Distortion scale gauges the pattern of respondent answers for positive impression on the assessment.

Employees who are well matched to their position tend to have better work attendance, less turnover, higher job satisfaction, and superior job performance. The employee and the employer share the benefits of enhanced person-job fit; the customer also benefits from interacting with a top-performing agent of the company.

As part of the PXT Select™ line of products for employers, the CSP II is designed to facilitate achieving the best possible job fit. Our clients use the CSP II to develop effective work teams and design and implement optimal customer service. It also helps supervisors determine approaches that facilitate working with a particular employee and help develop training plans for customer service.

The CSP II is the product of collaboration between several test development psychologists and specialists. The instrument investigates three areas (Behavioral Traits, Proficiencies, and the Company Service Perspective) as a part of evaluating how an individual fits into a particular job. By reviewing broad areas of service perspective, behavioral characteristics, and basic proficiencies, we expand our ability to discover which

areas will be most effective in determining job fit for a given position.

Using Performance Models

The job match process for the CSP II represents an effective approach that minimizes the time required to efficiently describe jobs, people, and their degree of match. The descriptive process starts by comparing the score pattern of those who are most successful with those who are least successful within a specific position. The scores on each scale of the CSP II are reported on a ten-point scale. The higher a score on the scale, the greater the probability of observing the behaviors typically associated with the scale. When the scores of top and bottom performers tend to lie in unique ranges within the scale, we are able to define a model for that sample.

The resulting pattern of scores across the various dimensions measured by the CSP II serves as the initial Performance Model upon which the job matching is based.

The CSP II allows for the concurrent study of those incumbents available, job requirement assessments by those who know the job, or by both of these. Even with a small sample, you have a good place to start the process and an approach that will allow you to further refine the initial model with ongoing evaluation. It is important for a business to continually maintain the job relevance of the process by which employees are

placed. Once developed, any Performance Model may be continuously updated as more empirical information becomes available.

Based on the pattern that reflects what is related to successful service in a position, a Performance Model for each factor (each scale on the assessment) may be built. This Performance Model consists of a range along each scale in which the scores of the most effective performers tend to fall. These scores differ as much as possible from the range of scores typical of the least effective performers. Because the characteristics measured tend to vary among most people, the typical model will be three to five units wide.

The farther outside this range (Performance Model) one's scores fall, the less likely the individual will be a good fit to the job in regard to that particular characteristic; it takes a consistent pattern of poor fit to result in a poor overall match. In the reports, job match is reported as an Overall Percent Match to a specific Performance Model for a specific job. This allows for a variation between persons who may still share a good job fit to a position.

By using a Performance Model it is easier to quickly identify individuals who will fit well in a specific position. Additionally, the model allows one to identify if a respondent might have adjustments to make in order to perform more effectively. This information is important for both job placement and job training.

The range of overall percent matches reported will be from 25% to 95%. An overall percent match summarizes the results of the six Behavioral Traits scales plus the two Proficiencies scales. The higher the Overall Percent Match reported, the higher the expectancy the individual will fit well into the job under consideration. An additional test section, known as the Company Service Perspective, is matched to your company's unique stance on service and is not included in the Overall Percent Match.

Because the interpretations generated by the CSP II are provided in reports written in common business language, they require no psychological interpretation by the user. This allows the CSP II to be used by those with no special psychological training in practical situations common to the selection processes found in successful companies.

Note: While the Performance Model approach of matching individuals to a job provides information of great value and can be an informative part of the placement decision, the results from any selection assessment should never make up more than a third of the final decision-making process.

Chapter 2: Behavioral Traits

Initial Research Methodology

Our first step was to compare whether tasks or job requirements from an earlier validity study were significantly similar to the tasks or satisfactory job requirements identified from a major job analysis study performed by the Gallup Corporation. Our study was comprised of 3,000+ participants from over 1,000 jobs whereas the Gallup Corporation study for MCI targeted their Customer Service representatives.

Note: The original validity study, prepared by Dr. David Pearson and accepted as meeting all of the requirements for demonstrating validity as set forth in the Federal Uniform Guidelines on Employee Selection Procedures, referenced as 29 CFR 1607 (1979), was for a major international manufacturing corporation.

The validation study for the major corporation with over 3,000 employee participants consisted of a job analysis where employees were asked to list five to ten of the most important tasks required for satisfactory job performance.

From these task statements, key words were identified and computer-compiled. A frequency distribution of key phrases was then prepared as the result of a factor analysis statistical program. Similarity of key phrases across 1,000 job titles was examined.

Jobs with a high degree of overlap with respect to the employees' task statements were clustered into job families. Among these were task requirements identified as customer satisfaction factors. The high frequency key phrases for each job family were then expanded back into sentences considered appropriate for use in a performance measurement instrument.

Verification and overall agreement of the identified job groups as to key phrases and use of words were presented to all supervisors and management personnel involved in the construct validation study. Eight trait factors, or attitudinal indexes, were established for an overall job family entitled Customer Service. That is, there was a large clustering of task phrases that indicated there were similar job requirements for what was generically called Customer Service factors or traits. The CSP II uses six of these eight factors.

Thus, eight trait factors, or attitudinal indexes, that correlated to a significant degree with those formerly identified by the Gallup Corporation in an in-depth job analysis study and subsequently corroborated with MCI's managers, and those in a position to know, i.e., immediate supervisors of the employees, were incorporated into the Customer Service Perspective, the predecessor of the current CSP II. These factors were designated with appropriately named descriptions as listed in Table 2.1.

Table 2.1

CSP Behavioral Traits scales

Trust	Conformity
Tact	Focus
Empathy	Courtesy
Conscientiousness	Flexibility

To help meet the standard of minimizing the time required for a respondent to complete the assessment, it was determined that ten questions per trait, factor, or attribute would be sufficient to "tap" the strength and/or weakness of an individual's attitude toward particular job requirements. These traits were identified as being important based upon the customer service job description. The final questions and trait factors were developed after extensive discussions with supervisors and others at MCI.

The questions for each trait were written to eliminate hostility on the part of the respondent and, at the same time, preclude issues of invasion of privacy, abridgment of one's right against self-incrimination, or invoking blatant or implied discrimination. Words that might be offensive to any particular ethnic group or gender, ADA-relevant status, and other relevant considerations were not used. As to writing questions considered fair or non-hostile, it was obvious that if a respondent reacted negatively to a question, he or she may provide an inaccurate answer, if they answered at all, out of consternation or confusion. However, the questions were purposefully written and arranged in a manner

that provided information as to whether or not the results were distorted.

Once the Customer Service Perspective questions were reviewed as acceptable within the parameters discussed earlier, the questions were assigned a random number and placed in their assigned order. Details concerning the statistical evidence of the effectiveness of the final set of items are covered later in this chapter (for Behavioral Characteristics) and Chapter 3 (for the Proficiencies section).

Finally, while models utilized in the CSP II to gauge top performers may vary across each scale, we expect a number of the models developed will tend toward the high end of the scale. This is because each scale measures constructs integral to those responsibilities common to employees who deal directly with customers. However, while we describe many of the following traits in terms of high scores and high performance, it is conceivable the results of any particular company's Overall Job Match process may yield models toward the low end of the scale when their unique performance goals and vision for excellent service call for it.

Understanding the Factors Measured by the CSP II

While the initial research yielded eight factors associated with customer service, we conducted studies that lead to the creation of the CSP II, including our

distillation of these factors to the six most influential traits related to customer service. These six Behavioral Traits are defined below. The *Remarks* paragraphs detail the underlying principles associated with including each factor in a customer service assessment. However, while each remark provides logical reasons why high scoring results may reflect better customer service, model development in any particular company may reasonably and accurately conclude that lower scores on a specific scale lead to better performance in light of the particular service style of that company.

TRUST

Definition: Measures the degree to which one accepts and has confidence in what others say and do. Often unaffected by, or perhaps even unaware of, criticism or cruel acts of others; will often convey kindness and compassion despite such unpleasantness. A trusting person is one who does not automatically distrust the motives of others or react in a skeptical manner with the resultant behavior often producing a negative and unpleasant demeanor.

Remarks: A customer must be able to rely upon and trust the motives, advice, or remarks of others, particularly sales and service personnel. A trusting employee is more likely to instill the atmosphere around them with confidence, optimism, assurance, etc., which in turn affects everyone's performance positively. With a higher degree of trust, one may find it easier to deal with a normal situation in a timely or positive manner.

TACT

Definition: Measures one's ability to deal with confrontations or unpleasantness in a constructive manner. One who deflates anger or harshness from others with non-pejorative and non-inflammatory remarks. A diplomatic person is one who really strives to interact with others in a calm, soothing, pleasantly modulated voice despite the occasional temptation to confront an obnoxious and rude person inappropriately.

Remarks: One's mannerisms or remarks to a customer will surely influence the customer toward or away from future purchases or service requests and may result in a verbal or written complaint to management if conducted poorly. Losing customers–some of whom give fair warning, some of whom never come back–is detrimental to a company's business future. Having the willingness to choose a diplomatic path with a customer is sure to enhance the sale or service.

EMPATHY

Definition: Measures the willingness to consider the unique perspectives of others, i.e., being eager to relate to the other person's point of view or concerns in an agreeable or tolerant fashion.

Remarks: Simply put, if one is considerate, he or she finds it easy, if not enjoyable, to understand another person's situation, thoughts and feelings, (e.g., recognizes, anticipates, and satisfies internal or external customers' needs). Such an individual will most likely

portray a genuine image of being tolerant, intuitive, and easy to deal with.

It should be noted that in some related questions, it is not asked if the respondent likes an unpleasant person but whether they can deal with an unpleasant person. If one genuinely wants to deal with some customers who may be unpleasant at times, one's attitude will quite often be conveyed in a pleasant manner of understanding and empathy.

CONFORMITY

Definition: Measures how comfortable one may be with following the rules and regulations of the workplace, organization, or institution to which they belong. A conformist does not question and will readily and willingly submit to others in authority. A conformist is not likely to be overly aggressive toward authority. He or she will more likely follow instructions and will follow through in a manner and timeframe specifically requested of him or her.

Remarks: Being able and willing to follow established policies and procedures is not easy for those who prefer to exert more autonomy than possible in a given organization. Taking and accepting orders, advice, or instructions seems to be agonizingly difficult for a fair number of individuals. In some cases, those who resist authority may question a situation, even when what is suggested makes perfectly good sense. Of course, others with more individuality and independence may offer fresh ideas to a customer service situation.

FOCUS

Definition: Measures one's ability to stay focused and to concentrate on the matter at hand with respect to the requirements of a task, despite interruptions or distractions. Included in this factor is the ability to return to one's thought processes and immediate goals at the point of the interruption.

Remarks: A person who continually operates equipment and/or who is required to keep accurate records is often expected to pay close attention to what he or she is doing at all times in service of the customer's needs. A customer can be quite reactive–often rather negatively–when discovering items or services missing or not included in their service experience.

The above sounds obvious, but all too often a person is hired who cannot stay focused upon a task at hand due to his or her vulnerability to being distracted.

FLEXIBILITY

Definition: Measures one's willingness to try, implement, or accept new ways of doing things and thus entertain new ideas or suggestions from a supervisor or leader in an amenable and cooperative manner. One who is not restricted by the routine of doing things that need to be revised.

Remarks: Change can be a frightening thing to some employees, thus creating negative consequences

when management implements new approaches to enhance productivity and service. An employee who is open to change can often create a progressive attitude that encourages a smooth change process. Today, a company or corporation often needs flexibly minded employees in order to meet the ever-changing consumer-based demand for new or better products and services. Those employees who are adaptable are often indispensable.

In summary, Table 2.2 aids in identifying each scale's unique contribution to the CSP II.

Table 2.2

CSP II Behavioral Traits scales

Low Scores	Scale Descriptions	High Scores
Wary Vigilant Skeptical	Trust: Tendency to believe the motives of others are reasonable and beneficial to oneself	Unquestioning Uncritical Optimistic
Direct Obvious Forthright	Tact: Ability to state a position without offending others	Discreet Diplomatic Restrained
Detached Indifferent Distant	Empathy: Capacity to understand another's situation and feelings	Understanding Compassionate Sensitive
Inventive Free-spirited Independent	Conformity: Tendency to comply with the rules and respect those in authority	Respects authority Comfortable with procedures Conventional
Distractible Preoccupied Inefficient	Focus: Ability to pursue an objective regardless of distractions	Attentive Purposeful Efficient
Uncompromising Inflexible Cautious	Flexibility: Tendency to explore new approaches to doing things; open to change	Adaptable Accepting Enjoys new approaches

Descriptive Statistics

The original CSP was analyzed with respect to certain requirements that must be met in order for the assessment to be useful for practical purposes. The first and most basic of these is that the assessment effectively distinguishes one examinee from another. An assessment that assigns all examinees the same score is obviously useless. An effective assessment distributes the scores of examinees across each scale in a meaningful manner.

Two kinds of statistics are helpful in determining whether this is the case. The first type measures the location, or central tendency, of the score distribution. The mean is such a statistic. It measures the average of the score distribution range. The second set of statistics indicates the extent to which scores are spread around the mean of the distribution and across the score range. The standard deviation, minimum score, maximum score, and standard error all characterize the *shape* of the score distribution, the ability of the assessment to distinguish examinees, and the degree to which their scores are considered normal or significantly different to the norm.

In 2003, 275 customer service professionals, sales professionals, and other professionals were administered the original Customer Service Perspective. This sample represented a mix of gender as well as a diverse range of educational and cultural backgrounds. The analysis results derived from the sample are

summarized in Table 2.3, which contains descriptive statistics for the original eight CSP Behavioral Traits scales.

Table 2.3

Behavioral Traits, descriptive statistics: the CSP, 2003 (N=275)

Scales	Mean	Median	Mode	Standard Deviation	Minimum	Maximum	Standard Error
Trust	6.2	7	7	2.22	0	10	.13
Tact	6.7	7	8	2.28	0	10	.14
Empathy	7.5	8	8	1.50	3	10	.09
Conscientiousness*	8.3	9	9	1.46	2	10	.09
Conformity	4.6	5	6	2.20	0	10	.14
Focus	7.7	8	10	2.35	0	10	.14
Courtesy*	7.8	8	8	1.09	3	10	.06
Flexibility	8.2	9	10	1.88	1	10	.11

* No longer measured by the CSP II

As can be seen in Table 2.3, these scales had distributions of scores that tended toward the high end of each scale. The sample included professionals actively engaged in responsibilities related to servicing customers, from customer service representatives to salespeople and managerial staff.

Development of the Distortion Scale

In addition to the eight personality scales described in Table 2.2, the original CSP included a Distortion scale, which is also included in the current version. The Distortion scale was designed to identify unusual response patterns that could suggest a distortion of the results. This scale consisted of ten questions and was developed rationally using questions endorsed consistently *Yes* or *No* by a sample of 214 job applicants.

For example, 96.7% of this sample responded "Yes" to the following question, “Have you ever had a negative thought about another person?” It can be inferred that an applicant who responds, “No, I have never had a negative thought about another person,” could be attempting to create a less than representative positive impression as the overwhelming majority of individuals in our culture engage in the commonplace, albeit less than perfect, behavior in question. Therefore, an “uncommon” response to any one of these ten questions is an *infrequent* occurrence. Consequently, a consistent pattern of “uncommon” responses may lead one to question the respondent’s mindset when taking the test. It is possible misrepresentation is occurring

when multiple responses suggest characteristics better than reality often suggests.

The greater the number of questions answered in the infrequent direction, the greater the concern about how forthright the individual might have been. The implication is that a consistent pattern of unusually positive self-representation may infer less-than-forthright responses to the other questions of the assessment.

Table 2.4 indicates the percentage of the total sample that responded "yes" or "no" for each question in the Distortion scale of the CSP.

Table 2.4

Percentage of endorsement for each Distortion question (N=214)

Distortion Questions	Percent	
	Non-distortion response	Distortion response
First Question	96.7	3.3
Second Question	96.7	3.3
Third Question	85.5	14.5
Fourth Question	88.3	11.7
Fifth Question	96.7	3.3
Sixth Question	93.5	6.5
Seventh Question	91.1	8.9
Eighth Question	95.3	4.7
Ninth Question	71.5	28.5
Tenth Question	77.1	22.9

According to Table 2.4, the range of Distortion responses for the ten questions was from a low of 3.3% to a high of 28.5% for a sample of 214 job applicants. Additionally, in Table 2.5, we can see that even if an individual identifies with a few items on this scale, it takes a consistent declaration of positive self-impression (considered rare as evidenced by the data) to suggest that a response pattern is questionable (receiving a score of note on this scale).

Table 2.5

Accumulation of Distortion endorsements (N=214)

# of Distortion items endorsed	Frequency	Percent	Cumulative percent
10	0	.0	.0
9	0	.0	.0
8	0	.0	.0
7	0	.0	.0
6	0	.0	.0
5	6	2.8	2.8
4	2	.9	3.7
3	5	2.3	6.1
2	16	7.5	13.6
1	35	16.4	29.9
0	150	70.1	100.0
Total	214	100.0	

Thus, it is apparent that approximately 94% of the sample responded to three or fewer of the Distortion items in a fashion suggesting positive self-representation. In

addition, scores that signify an endorsement frequency of six or more Distortion items is unheard of in this sample. This suggests that such a response pattern in any future respondent is cause for considering that his or her other responses (those in the Behavioral Traits section of the assessment) may not accurately represent his or her traits or attitudes.

That is, the respondent's answers may suggest a more positive-than-actual representation. Determination of exceptionally high distortion item selection will be noted in the report. See your CSP II Report Guide for details on Distortion reporting.

Reliability Analysis

No measure can be of much value unless it measures traits in a reliable or consistent manner. Reliability refers to the consistency of assessment items (questions in a scale) as selected by respondents, but it can also concern the scores obtained when retested with the same assessment on different occasions. Because all types of reliability are concerned with the degree of consistency of test items and their sums as reflected in scores over time, they are expressed in terms of a correlation coefficient.

A correlation coefficient expresses the degree of relationship between two variables. This relationship is expressed as a decimal number ranging from .00 to 1.00, where 1.00 indicates absolute consistency and .00 indicates the absence of any consistency. Although no test is a perfectly reliable instrument, test reliability

correlation coefficients (Cronbach alphas) should be in the .70s, .80s, and .90s (U. S. Department of Labor, 1999), while test-retest correlations should manage as high a relationship between administrations of the test as possible (more on this process later).

Coefficient Alpha Reliability Analysis

Coefficient alpha indicates the consistency of responses to individual test questions. The higher an assessment's coefficient alpha, the more consistent the questions are for that test. An assessment with a low coefficient alpha could produce different or inconsistent scores each time the same person takes the assessment. A high coefficient alpha, however, indicates the test has a consistent response rate by the respondents in the sample studied in a single administration of the test and thus is more likely to derive similar results in future administrations.

Edward J. Inderrieden, Ph.D., Professor in the College of Business Administration at Marquette University in Milwaukee, Wisconsin, performed a Cronbach alpha reliability analysis (a model of internal consistency based on the average inter-item correlation) on the questions contained in each of the eight scales of the original Customer Service Perspective. This was accomplished with the results indicating an alpha of .75 or better for each scale. Subsequent studies, focusing on the assessment's current form of the CSP II are reviewed in Chapter 5 and are also supportive of the reliability of the scales.

Test-retest Reliability Analysis

When analyzing the reliability of an assessment, it is often relevant to observe the consistency of test scores as time passes. By re-administering the assessment after an appropriate amount of time (typically at least a few weeks up to about six months), individuals' scores may be compared between each administration. The assumption here is that a consistent set of questions should yield approximately the same scores if taken later. To do this, a correlation of scores from one administration to the next is conducted. The higher the correlation coefficient, the more reliable the test's scores are seen to be.

Because the concept of test-retest reliability is easily understood, business users tend to appreciate seeing this in technical reports. However, as the literature points out, the test-retest reliability scores, when considering behavioral traits, tend to be lower than for other forms of reliability as they measure consistency of assessment items indirectly through raw scores.

Table 2.6 shows test-retest correlations for the eight original behavioral scales for a sample of job applicants and incumbents. Calculating test-retest reliability determines how consistent scores may be for individuals who take the test again over a period of time; four weeks in this case. This analysis indicates the eight original behavioral scales are reliable and produce consistent results as demonstrated by the high correlation coefficients listed below. Correlations of .32

or greater (whether positive or negative) are considered noteworthy (significant at the .01 level) for a sample of 61.

Table 2.6

Test-retest correlations, CSP (N=61)

Original CSP Scales	Test-Retest Correlations
Trust	.88**
Tact	.91**
Empathy	.61**
Conscientiousness	.75**
Conformity	.90**
Focus	.83**
Courtesy	.69**
Flexibility	.82**
Average	.80

** sig. at .01 level

According to Table 2.6, average test-retest correlation was .80, ranging from a low of .61 for the Empathy scale to a high of .91 for the Tact scale. This demonstrates high consistency of scoring over the period specified. There was a four-week interval between test administrations.

Validity

The software used with the CSP II allows the user to establish individualized and appropriate standards for establishing the level of performance for their employees. Once selected by the company, these top performers become the model (criterion) for the company's individualized use with the CSP II. The

criterion, or performance rating, is a measure that is essential in validity studies.

Validity and the Customer Service Perspective

Validity measures the extent to which an assessment evaluates what it is supposed to evaluate. In other words, it is the effectiveness of an assessment to predict important factors like performance or identify individual behavior. The concept of validity refers to the appropriateness, or accuracy, of inferences or decisions made about individuals based on test results. It is important from both a business and legal perspective that organizations answer the question regarding the validity of the inference underlying selection testing. If the inference is not valid, organizations may waste their business investment in selection instruments and risk challenges to their hiring and placement decisions.

Test validation refers to the process of gathering evidence to support the inference being made. Various approaches to gathering validity evidence will be described below.

The Construct Validity Approach

Construct validity refers to the extent to which a psychological measure is an accurate measure of a particular construct. Construct validity addresses the following questions, "What does this measure and how well does it measure what it was designed and purported to measure?"

Although there are different methods for evaluating construct validity, one informative method is to examine the correlation coefficients among the scales within the instrument. The pattern of correlations should be consistent with expectations based on what these scales were intended to measure. Scales that measure related personality constructs should converge, whereas scales that measure unrelated constructs should not correlate at all.

Incidentally, scales that are understood to represent traits that are typically in opposition to one another (hypothetically, like such traits as cooperativeness and competitiveness) should result in a correlation coefficient that is negative, signifying how high levels in one trait equate to low levels in the other. This is as informative a correlation as one that happens to be positive.

Correlations Between the Original CSP Scales

The original Customer Service Perspective was designed to measure eight primary dimensions of customer service related performance. Table 2.7 contains the correlations among the eight CSP scales for 275 customer service, sales, and other professionals in our sample.

Table 2.7

Original CSP scale inter-correlations (N=275)

CSP Scales	Trust	Tact	Empa.	Consc.	Conf.	Focus	Cour.	Flex.
Trust	1.00	.18**	.15*	.05	.14*	.03	.19**	-.05
Tact		1.00	.25**	.00	.40**	-.01	.35**	-.18**
Empathy			1.00	.07	.17**	.21**	.24**	.11
Conscientiousness				1.00	.06	.08	-.06	.07
Conformity					1.00	.06	.16**	-.46**
Focus						1.00	.07	.16**
Courtesy							1.00	-.05
Flexibility								1.00

** significant at the .01 level * significant at the .05 level

According to Table 2.7, it is evident that half of the correlations (14/28) between scales were so low as to suggest no relationship between each pair of scales, evidence of the uniqueness and validity of the original CSP scales. This implies each of these scales measured separate constructs or traits.

Correlations that range between approximately .14 and .25 (positive or negative) suggest a degree of relativity while maintaining separate and unique constructs (11 more of the 28 correlations). For example, Tact and Empathy were related with a correlation of .25. This follows conventional wisdom as these separate traits are seen as logically related while the correlation was not so high as to suggest the scales were identical. Discussions on basic construct development demonstrate how these scales measured different traits, while this correlation suggested they shared some degree of

relatedness. Even so, separate measurement of each scale provided the user with discrete construct data that described much about the respondent, especially when the scale scores were not similar in a particular case. This happens often with a correlation at this level.

Finally, the last three of the 28 correlations in Table 2.7 beyond .25 (positive or negative), were associated with scales much like those discussed in the previous paragraph. Yet, the probability of a respondent achieving dissimilar scores on two such scales was even more unlikely depending on the individual correlation coefficient found in the table. Even more so than before, when the respondent does indeed score dissimilarly in two such scales, one learns much in regard to their fit for a particular Performance Model (see the CSP II Report Guide).

The scales concerned are Tact and Conformity, Tact and Courtesy, and Conformity and Flexibility, but in the case of Conformity and Flexibility, the correlation is inverse, which means the higher a score on one scale, the lower the score on the other scale. All three of these pairs seemed to bear logical scrutiny when considering possible reasons for their correlations, which supported the idea that such traits related to one another one way or the other. The relativity of the current CSP II scales is discussed in Chapter 5.

The Criterion-related Validity Approach

Criterion-related validity is one of the most accepted means of demonstrating assessment validity for all types of assessments including general mental ability tests or tests that measure abstract reasoning. In a criterion-related validation study, scores on the target assessment may be correlated with job performance scores (the criterion) for the same group of applicants or employees. The correlation between assessment scores and job performance scores reflects the criterion-related validity of the assessment. The stronger the correlation, the more confidence one may have that the assessment is working as intended to select applicants based on a performance index.

Research Designs Used for Criterion-related Validation

There are two basic research designs used to conduct a criterion-related validation study. The difference between the two designs is the amount of time allowed to elapse between collecting the predictive data and the criterion data. Using the concurrent validation method for demonstrating criterion-related validity, the predictive data (e.g., assessment scores) and the criterion data (e.g., job performance ratings) are collected at approximately the same time. Using the predictive validation method, the predictive data (e.g., assessment scores) are collected and a certain amount of time is allowed to elapse before the criterion data (e.g., job performance ratings) are collected. As opposed

to the concurrent validation method, the predictive validation method is longitudinal in nature.

Job Match Percentage Criterion-related Validity Study

This section of the Research Report reviews the statistical relationship between the Overall Job Match Percentage (OJMP) and a performance criterion. This criterion was based on an executive committee evaluation of each member of the sample. Sixty three employees were administered the original CSP. Those who completed the assessment were rated by the executive committee. Ratings included top, middle, and bottom performance groups. The sample included 63 members of a sales and service company with clients nationwide. Table 2.8 displays the average scores and standard deviations for the sample, including their scores, ratings, and job match percentages.

Table 2.8

Job Match Percent validation: descriptive statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Performance Rating	63	1.00	3.00	1.9524	.8118
Trust	63	1.00	10.00	6.5079	2.3819
Tact	63	1.00	10.00	6.9683	2.1400
Empathy	63	4.00	10.00	7.5397	1.2678
Conscientiousness*	63	2.00	10.00	8.3810	1.5804
Conformity	63	1.00	10.00	5.4603	2.3883
Focus	63	1.00	10.00	8.2222	2.2248
Courtesy*	63	5.00	10.00	7.8889	1.1375
Flexibility	63	2.00	10.00	7.8095	1.9745
Vocabulary	63	8.00	10.00	9.7460	.5948
Numerical	63	7.00	10.00	8.9841	.9068
Distortion	63	5.00	10.00	8.5397	1.2290
Overall Job Match %	63	65.00	95.00	84.6349	8.7239

*Scale no longer included in the CSP

As is evident from these data, the performance groups were evenly distributed while the sample's scale scores tended toward the high end to some degree. With this data and the original scores/ratings/OJMP for each individual in the study, statistical analysis may reveal the relationships between the rankings and the OJMPs for the sample. If the CSP is measuring what it is supposed to measure (has construct validity) and predicts what is vital to identifying top performers (has criterion-related validity), then the relationship between rankings and OJMPs should be high. In other words, correlations should appear quite high between these two factors if the OJMP accurately selects top performers as set by

the criterion. Table 2.9 demonstrates the correlation coefficient that results from comparing the OJMP and the performance rating (criterion).

Table 2.9

OJMP and criterion correlations

N = 63	Overall Job Match	
Performance Rating (Criterion)	Correlation Coefficient Sig. (2-tailed)	.90** .000

** Correlation is significant at the .01 level (2-tailed).

The correlation coefficient of .90 signifies an exceptionally high relationship between the predictor variable (the OJMP produced from CSP model development) and the criterion (performance rating). This establishes the validity of the matching process and the predictive strength of the CSP. Note that the criterion is not a measure of test scores nor is it in any way related to actual scores. Such a criterion would spuriously yield high correlations. Rather, the criterion is related solely on performance as determined by the aforementioned committee; therefore, the correlation between the performance ranking and the OJMP passes the statistician's primary counterfeit-criterion test well.

While we are discussing criterion-related validity, this study is a good example of how the job match process is so much more effective than a straight linear way of thinking about scale scores and predictive validity. If an

assessment were utilized in which scale scores alone were determinate for predicting success, one may find (and often does) traditional thinking about test scores falls short.

Higher is not necessarily better in the case of assessment scores. As evidence to this statement, we can observe the score distribution of the sample based on the groups identified by performance ratings. In most cases, we have found high scores do not neatly equate to high performance ratings. Rather, there is a range of scores common to the high performance group that lies somewhere along the ten-point STEN scale, sometimes below the highest range of possible scores. This is, of course, dependent on the specific sample of an organization.

When the score distribution of the top performers differs from that of the bottom performers, we can then determine predicted success. This is what the OJMP measures. It is a percentage score that evaluates the scores of an individual in comparison to the range of scores common to top performers. Since the top performers' score distribution is not necessarily in the high score range of the scale, attaining a high score may have a detrimental effect on one's OJMP. This illustrates how higher is not necessarily better in the case of predicting success. Table 2.10 shows how correlating a scale distribution does not yield high marks when compared to performance ranking. This is a direct result of the fact that the OJMP is not based simply upon high score attainment but also on the

range of scores typical of the top performing group based on a company's specific criteria.

Table 2.10

Correlations between OJMP and scale scores (N=63)

	r	p
Trust	.039	.763
Tact	.096	.452
Empathy	.018	.888
Conscientiousness*	.059	.646
Conformity	.089	.489
Focus	.019	.882
Courtesy*	-.070	.584
Flexibility	-.200	.116
Vocabulary	-.029	.820
Numerical	.055	.668

*Scale no longer included in the CSP

These correlation coefficients demonstrate how it is often the case that top performers' scale scores will not relate to a linear, "higher is better" way of thinking. If that were the case, correlations would reflect the situation of high scale scores equaling higher OJMPs. That's the efficacy of Performance Modeling; taking into account that top performers do not necessarily follow the erroneous, conventionally accepted tendency to score highly on assessment scales. For more on this topic, please refer to Appendix B of this report.

Scale Scores Criterion-related Validity Study

Returning to the discussion of the assessment development process, we will review the results of analyzing the scores of the sample group from MCI with performance ratings to determine the criterion-related validity of the original CSP.

Test booklets and one-page answer sheets were printed for distribution to four widely geographically separated MCI Customer Service centers.

The distribution of the questions and answer sheets was accomplished by the manager in charge of the MCI research study who chose the following Customer Service centers as participants for this research project:

- Atlanta, Georgia
- Cedar Rapids, Iowa
- Pinellas Park, Florida
- Sacramento, California

The initial participant selection included 50 participants from each of the four centers listed above. Although the study was voluntary, there were many who chose not to participate because of personal reasons and/or because they did not want their names used in MCI's ranking scheme, regardless of the fact their names would be sent only to the researchers for validation purposes and not otherwise disclosed.

Participants were rated in only two very broad categories of either A's (top performers) or B's (bottom performers).

Now, the problem of ranking people into only two broad categories rests on the fact that some of those rated high may be barely above the low or midpoint cut-off of high versus low while others may be about half-way up the high side and still others at or near the very top of the high side. The same problem would adhere to ranking people on the low side of a midpoint i.e., some people rated as low may be just below the high side of a midpoint cut-off, etc.

Consequently, from a statistical normal curve standpoint, the scores of all those in the middle range (i.e., those considered high but not too high and those considered low but not too low) can "wash out" any significant score differences that may otherwise be found to exist for those considered.

The correctness of those rated, as well as the fact that respondents were not allowed to take the questionnaire home where someone else could contribute to their responses, was verified by the researchers.

Once all answer sheets were scored and the data verified, frequency distribution tables were developed and analyzed as shown below.

The inconsistency of ratings is a very common occurrence and often difficult to correct or avoid. However, in this case, nearly everyone, or at least a

majority in a position to know within MCI's customer service center organization, not only agreed upon what behavioral attributes or traits were needed or wanted according to their uniformly used job description requirements, but they also used the rating system therein in a fairly consistent manner. This reduced the probability for error considerably. The data in Tables 2.11 (A&B) confirm this assumption.

Table 2.11A

MCI - Frequency distribution - profile of those rated as TOP *N=24*

Trait Names, Original CSP	0	1	2	3	4	5	6	7	8	9	10	MEAN	STD.
Conformity	0	0	0	1	4	8	7	2	1	0	1	5.54	1.13
Conscientiousness*	0	0	0	0	0	0	5	7	8	4	0	7.46	0.97
Courtesy*	0	0	0	0	0	5	9	8	1	0	1	6.38	0.83
Empathy	0	0	0	0	0	3	3	12	6	0	0	6.88	0.92
Flexibility	0	0	0	0	0	0	4	7	9	4	0	8.04	0.99
Focus	0	0	0	0	0	0	2	6	13	3	0	7.71	0.72
Tact	0	0	0	0	1	2	3	9	4	5	0	7.17	1.32
Trust	0	0	0	4	4	3	2	6	3	2	0	5.79	2.04

*Scale no longer included in the CSP
Highlighted area broadly demonstrates mean and standard deviation

Table 2.11B

MCI – Frequency distribution – profile of those rated as BOTTOM N=26

Trait Names, Original CSP	0	1	2	3	4	5	6	7	8	9	10	MEAN	STD.
Conformity	1	0	6	1	6	7	4	1	0	0	0	4.04	1.68
Conscientiousness*	0	0	0	2	5	7	6	5	1	1	0	5.38	1.30
Courtesy*	0	1	1	7	5	4	7	1	0	0	0	4.35	1.49
Empathy	0	0	0	2	4	10	8	1	1	0	0	5.19	1.11
Flexibility	0	0	0	1	4	5	8	4	4	0	0	5.85	1.17
Focus	0	0	0	2	2	8	5	6	2	1	0	5.81	1.47
Tact	0	0	1	5	6	4	4	2	3	1	0	5.08	1.86
Trust	0	6	8	3	3	6	1	0	0	0	0	3.04	1.60

*Scale no longer included in the CSP
Highlighted area broadly demonstrates mean and standard deviation

Table 2.12 depicts the statistical analysis results for each of the eight Traits names for the TOP group versus the BOTTOM group in terms of the mean differences and the level of significance found to be present between these two groups.

Table 2.12

Score differences by performance group

Trait Names	Mean (Top Group) N=24	Mean (Bottom Group) N=26	T-value	Level of Significance
Conformity	5.54 (1.13)	4.04 (1.68)	3.10	.01
Conscientiousness*	7.46 (0.97)	5.38 (1.30)	5.93	.001
Courtesy*	6.38 (0.83)	4.35 (1.49)	5.08	.001
Empathy	6.88 (0.92)	5.19 (1.11)	5.46	.054
Flexibility	8.04 (0.99)	5.85 (1.17)	3.92	.001
Focus	7.71 (0.72)	5.81 (1.47)	5.74	.001
Tact	7.17 (1.32)	5.08 (1.86)	4.23	.001
Trust	5.79 (2.04)	3.04 (1.60)	3.51	.001

*Scale no longer included in the CSP
Standard deviations are shown in parentheses

The major validation study conducted by Dr. David Pearson, upon which the traits and questions for the MCI study were based, involved a number of research steps:

(1) From the major study, at least 23 job groups were identified using factor analyses as the result of the elicitation technique of job requirement statements from approximately 3,000 employee participants. From the 23 job groups, a number of these were considered to belong to a family of jobs encompassing service-related tasks. There were eight

of these service-oriented factors, or traits, used in the study of MCI customer service personnel.

(2) The eight factors, or traits, were then presented to MCI's human resources staff for approval and verification that these eight factors indeed identified in a significant, majority manner, their concepts or ideas as to traits needed in their organization in order to meet customer service job performance criteria.

(3) In MCI's efforts to be thorough in their ideas as to what constitutes needed personality traits for their customer service personnel, they retained the nationally known and respected Gallup organization to perform an in-depth job or task analysis of their presently employed customer service employees of which there were at least 300 who participated.

(4) The results of Gallup's job or task analysis were then forwarded to the researchers. Comparisons were made as to similarities between MCI's job requirements (task statements) and those from the original validation study conducted earlier. This resulted in the final eight traits, or attitudinal factors, presented here as Conformity, Conscientiousness, Courtesy, Empathy, Flexibility, Focus, Tact, and Trust.

(5) The eight factors, or traits, and the questions used to "tap" the strengths or weaknesses of the eight factors or traits were then returned to MCI for final approval and confirmation that these comported with or identified MCI's job requirements for a customer service employee.

(6) After acquiring the data from the MCI study as iterated in Tables 2.11 A&B and Table 2.12, it was very clear that the eight factors, or Behavioral Traits, did indeed predict and identify those who would succeed as top performers and those whose scores revealed weaknesses that resulted in their being classified as bottom performers.

(7) The eight factors utilized by the original Customer Service Perspective were valid predictors when considering customer service employees designated most likely to succeed.

Gender Issues: Disparate Impact

A study was conducted by Dr. David Pearson to investigate the effects of gender on the results of the original eight CSP Behavioral Traits scales. The eight Behavioral Traits scales were a part of this study conducted with 300 MCI employees. The composition of this group was approximately 19% female and 81% male. For each of these scales, no significant differences in the results were found with regard to the gender of the participants.

An additional study was conducted with a sample of 214 subjects classified by gender for an analysis of variance that demonstrated some differences by gender. Note the differences in means when examining Table 2.13. These variations, while statistically different, did not demonstrate any "real-world" difference.

Table 2.13

Analysis of Variance (ANOVA) by gender, descriptives

CSP Scale	Gender	N	Mean (higher in bold)	Standard Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Trust	male	146	6.5479	1.9725	.1632	6.2253	6.8706	1.00	10.00
	female	68	**6.6471**	2.1494	.2607	6.1268	7.1673	1.00	10.00
	Total	214	6.5794	2.0258	.1385	6.3065	6.8524	1.00	10.00
Tact	male	146	6.4110	2.2914	.1896	6.0361	6.7858	.00	10.00
	female	68	**7.2647**	2.2302	.2704	6.7249	7.8045	1.00	10.00
	Total	214	6.6822	2.3017	.1573	6.3721	6.9924	.00	10.00
Empathy	male	146	7.3767	1.5092	.1249	7.1298	7.6236	3.00	10.00
	female	68	**7.5147**	1.3326	.1616	7.1921	7.8373	3.00	9.00
	Total	214	7.4206	1.4537	.0993	7.2247	7.6164	3.00	10.00
Conscient.*	male	146	**8.5548**	1.3444	.1113	8.3349	8.7747	2.00	10.00
	female	68	8.1618	1.5023	.1822	7.7981	8.5254	4.00	10.00
	Total	214	8.4299	1.4050	.0960	8.2406	8.6192	2.00	10.00
Conformity	male	146	4.0685	2.0262	.1677	3.7371	4.3999	.00	9.00
	female	68	**5.0882**	2.4234	.2939	4.5017	5.6748	.00	10.00
	Total	214	4.3925	2.2065	.1508	4.0952	4.6898	.00	10.00

Table 2.13 (continued)

Analysis of Variance (ANOVA) by gender, descriptives

CSP Scale	Gender	N	Mean (higher in bold)	Standard Deviation	Std. Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Focus	male	146	7.4932	2.3817	.1971	7.1036	7.8827	.00	10.00
	female	68	**7.5588**	2.5649	.3110	6.9380	8.1797	1.00	10.00
	Total	214	7.5140	2.4355	.1665	7.1858	7.8422	.00	10.00
Courtesy*	male	146	7.5822	1.0748	.0889	7.4064	7.7580	4.00	10.00
	female	68	**7.9853**	1.0147	.1231	7.7397	8.2309	5.00	10.00
	Total	214	7.7103	1.0704	.0731	7.5661	7.8545	4.00	10.00
Flexibility	male	146	**8.4247**	1.6768	.1388	8.1504	8.6989	2.00	10.00
	female	68	7.6471	2.3421	.2840	7.0801	8.2140	1.00	10.00
	Total	214	8.1776	1.9419	.1327	7.9159	8.4392	1.00	10.00

*Scale no longer included in the CSP

Table 2.13 demonstrates the basic descriptive statistics derived from this sample and lays the groundwork for the analysis of variance in Table 2.14. Differences between men and women in mean scores were as high as approximately one raw score point on the Conformity scale and as little as .07 of a raw score point on the Focus scale.

Table 2.14

Analysis of Variance (ANOVA) by gender

CSP Scale		Sum of Squares	df	Mean Square	F	Sig.
Trust	Between Groups	.456	1	.456	.111	.740
	Within Groups	873.694	212	4.121		
	Total	874.150	213			
Tact	Between Groups	33.815	1	33.815	6.549	.011
	Within Groups	1094.578	212	5.163		
	Total	1128.393	213			
Empathy	Between Groups	.883	1	.883	.417	.519
	Within Groups	449.266	212	2.119		
	Total	450.150	213			
Conscient.*	Between Groups	7.166	1	7.166	3.676	.057
	Within Groups	413.282	212	1.949		
	Total	420.449	213			
Conformity	Between Groups	48.242	1	48.242	10.343	.002
	Within Groups	988.786	212	4.664		
	Total	1037.028	213			
Focus	Between Groups	.200	1	.200	.034	.855
	Within Groups	1263.258	212	5.959		
	Total	1263.458	213			
Courtesy*	Between Groups	7.538	1	7.538	6.757	.010
	Within Groups	236.499	212	1.116		
	Total	244.037	213			
Flexibility	Between Groups	28.052	1	28.052	7.672	.006
	Within Groups	775.201	212	3.657		
	Total	803.252	213			

*Scale no longer included in the CSP

As illustrated in Table 2.14, the Tact, Conformity, Courtesy, and Flexibility scales yielded statistically significant differences between genders with the higher score trend going to women for Tact, Conformity, and Courtesy while men achieved the higher mean on Flexibility. Again, the actual differences in the means between these groups was not reasonably discernable on a practical level nor in practice as demonstrated when men and women were selected under the Performance Model system utilized in the CSP at that time. Despite what appears to be a statistical difference in mean scores, based predominately on differences of one raw score point or less, this does not create disparate impact based on gender thanks to the inclusive nature of the Performance Model development process. Nonetheless, this situation contributed to spurring our interest in creating more items per scale as explained in greater detail in Chapter 5.

Chapter 5 details the processes undertaken to increase the utility and quality of the CSP, including the creation of more questions per behavioral scale and the removal of two scales from the Behavioral Traits section.

Before discussing that process, the next chapter focuses on the same statistical issues reviewed so far as they pertain to the Proficiencies section of the CSP, both in its original and current forms, as these sections remained unchanged over time.

Chapter 3: Proficiencies

Overview of the Proficiencies Section

The Vocabulary and Numerical test items developed for the CSP were derived from corporate-employed Americans working in numerous job classifications and have been evaluated by university professors and certified as being at a junior high school level of comprehension (please see letter, Appendix A, from Dr. Clegg, professor at Texas Christian University, Fort Worth, Texas).

Using too few words or too few math problems – for example, five to ten – is typically an unfair and unreliable indicator of one's competence in these skills areas. On the other hand, too many test items – for example, 40 or more – at one testing session may be fatiguing and may result in a lower than normal level of performance.

There is recognized disparity between high schools, at nearly all grade levels, in terms of the depth of English grammar and the basic math being taught. That is, some students are exposed to an extremely sophisticated level of reading comprehension and/or math while others learn at a very elementary level.

Since poor reading comprehension, use of the English language, and the working of simple math problems are a much more common malady at all job levels than formerly realized, Dr. David Pearson–from his 30 years of experience with companies, corporations, and government entities–has developed this section of the

CSP for the purpose of clarifying this potential problem with a high degree of accuracy.

Dr. Pearson's design, originally entitled the Customer Service Perspective Basic Proficiencies section, is now called the Customer Service Profile II Proficiencies section and differs from the original in name only.

Independent Review

In order to acquire additional expert opinion concerning the Vocabulary section of the CSP during its initial development, Dr. Pearson sought the independent review of Luther B. Clegg, Professor and Chair, Department of Curriculum and Instruction at Texas Christian University. His observations support the utility of the Vocabulary section of the CSP and confirm, via independent review, the fairness of this test. See Appendix A for a copy of the letter.

Descriptive Statistics

One hundred and twenty-two customer service representatives and related professionals were administered the CSP Proficiencies section. These scores have been analyzed with respect to certain requirements that must be met in order for the test to be useful for practical purposes. The first and most basic of these is that the test effectively distinguishes one examinee from another. A test that assigns all examinees the same score is useless. An effective test's sample of scores distributes examinees across the score

scale in as even a way as possible based on forms of central tendency like the mean and standard deviation.

The mean measures location of the center of the score distribution, the "average" of the sample's individual scores. The second set of statistics indicates the extent to which scores are spread around the mean of the distribution and across the score range. The standard deviation, minimum score, maximum score, and standard error all characterize the "scatter" of the score distribution, or the ability of the test to distinguish examinees. Table 3.1 contains descriptive statistics for the two CSP Proficiencies scales.

Table 3.1

Proficiencies, descriptive statistics: CSP (N=122)

CSP Proficiencies Scales	Mean	Median	Mode	Standard Deviation	Min	Max	Standard Error
Vocabulary	17.57	18	18	2.17	10	20	.20
Numerical	16.76	18	18	2.99	3	20	.27

Table 3.2 summarizes the constructs measured by the CSP Proficiencies scales.

Table 3.2

CSP Proficiencies scales descriptions

Low Scores	Scale Descriptions	High Scores
Reflect a verbal insufficiency that may require some development	Vocabulary: understanding the meaning of words when used in sentences	Often associated with being proficient in basic language skills
Reflect a need for development of mathematical skills	Numerical: understanding basic mathematical concepts and working with numerical problems	Suggest a basic level of mathematical proficiency

Reliability Analysis

Reliability refers to the consistency of test items as selected by respondents as well as the scores obtained when re-tested with the same assessment on different occasions.

Coefficient Alpha Reliability Analysis

Coefficient alpha indicates the consistency of responses to individual test questions. The higher a test's coefficient alpha, the more consistent the questions are for that test. Table 3.3 contains coefficient alpha reliabilities for the two Proficiencies scales for a sample of 150 sales/customer service professionals. Internal consistency of the Proficiencies (Vocabulary and Numerical) was determined by calculating coefficient alpha reliability. This analysis indicates the Proficiencies scales are reliable and produce consistent results.

Table 3.3

Coefficient alpha, CSP Proficiencies (N=150)

CSP Proficiency Scales	Coefficient Alpha
Vocabulary	.96**
Numerical	.94**
Average	.95

** sig. at .01 level

According to Table 3.3, the average alpha coefficient was .95, with .94 for the Numerical scale and .96 for the Vocabulary scale.

Test-retest Reliability Analysis

The results of another study can be found in Table 3.4 with test-retest correlations for the two Proficiencies scales for a sample of working sales and service personnel. Test-retest analysis helps determine how consistent scores may be for individuals who take the test again over a period of time, four weeks in this case. This analysis indicates the two Proficiencies scales are reliable and produce consistent results, as demonstrated in the high correlation coefficients listed below. Correlations of .32 or greater are considered noteworthy (significant at the .01 level) for a sample of 61. The relatively small sample size allows smaller correlation to be practically and statistically significant, keeping in mind this is a different value than the commonly reported Cronbach's alpha coefficient.

Table 3.4

Test-retest correlations, CSP (N=61)

CSP Scales	Test-retest Correlations
Vocabulary	.68**
Numerical	.74**
Average	.71

** sig. at .01 level

Table 3.4 shows average test-retest correlation was .71, ranging from .68 for Vocabulary, to .74 for Numerical. This demonstrates high consistency of scoring over the period specified. There were four weeks between test administrations.

Validity

Validity is the extent to which an assessment measures what it is supposed to evaluate. Most of the validation studies for the Proficiencies section of the CSP were conducted concurrently with the Behavioral Traits section. These results may be found in Tables 2.8 and 2.10.

Intercorrelation of CSP Proficiencies Scales

The CSP was designed to measure two proficiencies related to customer service performance. Table 3.5 contains the intercorrelations among the CSP Proficiencies scales for 122 sales and service professionals.

Table 3.5

CSP Proficiencies intercorrelations (N=122)

	r	p
Vocabulary and Numerical	.36	.000

According to Table 3.5, high scores in Vocabulary tend to occur concurrently with high scores on the Numerical scale, though not at a rate suggesting equivalence of the constructs. The converse tends to be true as well with low scores on each. However, this correlation is not so high as to guarantee a certainty of this situation, suggesting the uniqueness of each scale. Obviously, different proficiencies are being measured, evident with only a cursory observation of the items in each scale. One assumption that may be made, based on this data, is that educational experiences of most individuals tend to emphasize both verbal as well as mathematical expertise, which can generate such a correlation in two distinct scales like these. Of greatest interest to the assessment-user are the cases in which one Proficiencies score far exceeds the other.

Gender: Adverse Impact Issues and the CSP Proficiencies Scales

A study was conducted to investigate the effects of gender on the results of the CSP Proficiencies section. The subjects represented a diversity of male and female sales and customer service professionals. The descriptive statistics from this study are found in Table 3.6, in which the *significantly* higher mean based on gender is listed in bold. Note that a higher mean does not imply that the mean is *significantly* higher, as is the case with the Vocabulary scale.

Table 3.6

Analysis of Variance descriptive, CSP Proficiencies

		N	Mean	Standard Deviation	Standard Error	95% Confidence Interval for Mean		Min	Max
						Lower Bound	Upper Bound		
Vocabulary	male	35	18.0571	1.7140	.2897	17.4684	18.6459	14.00	20.00
	female	26	18.0000	1.8762	.3679	17.2422	18.7578	14.00	20.00
	Total	61	18.0328	1.7698	.2266	17.5795	18.4861	14.00	20.00
Numerical	male	35	**18.0571**	1.8462	.3121	17.4230	18.6913	12.00	20.00
	female	26	15.8462	2.7084	.5312	14.7522	16.9401	10.00	20.00
	Total	61	17.1148	2.4906	.3189	16.4769	17.7526	10.00	20.00

Table 3.6 demonstrates the basic statistical data necessary for making analyses of variance. The first clear evidence is that Vocabulary means for each gender are virtually identical, while the difference between male and female Numerical means appears to be about two raw score points. After converting raw scores to STEN (standard ten-point) scores on the Proficiencies section, the resultant difference in STEN scores is about one point, clearly not enough to adversely impact individuals by gender in any practical manner.

The CSP utilizes a matching system with Performance Models that typically range from three to five or more STEN score points. As you can see below in Table 3.7, the difference in raw scores on the Numerical scale appears to be statistically significant, even if it is not practically different for the benchmarking system since the difference in means scores is only two raw points different.

Table 3.7

Analysis of Variance, CSP Proficiencies male/female

		Sum of Squares	df	Mean Square	F	Sig
Vocabulary	Between Groups	.0487	1	4.871E-02	.015	.902
	Within Groups	187.886	59	3.185		
	Total	187.934	60			
Numerical	Between Groups	72.926	1	72.926	14.377	.000
	Within Groups	299.270	59	5.072		
	Total	372.197	60			

Chapter 4: Company Service Perspective

In developing the CSP, our team of developers' experience with customer service assessments made evident the need for a customer service knowledge test that adapted to the needs of each individual company that utilized it. If every team of customer service professionals felt the same way about every customer service issue, a single answer key for the CSP would be sufficient. We have discovered that, while some concepts are universal among professionals, other issues are quite debatable and often reflect the individual style or perspective of an individual company. This is usually related to the company's vision and mission for their products and services and should definitely be tailored to fit the needs of that vision.

This sounds simple and logical except that modern testing practices tend to emphasize pre-set keys and norming tables in order for scoring algorithms to process results. Therefore, thinking outside the technologically limited box was mandatory, with a novel process for score tabulation being the result. Instead of gathering individual item responses and comparing them with a static key of correct answers, the data can be accumulated and compared with a unique key or response pattern developed for each individual company that uses the instrument.

Therefore, we are able to process a score that represents the alignment of responses between the respondent

and the company. This score signifies a conceptual fitness for the perspective held by the company concerning customer service issues. Thus, we have the utility of a matching instrument with the individuality of a corporate culture analysis: the Company Service Perspective section of the CSP. The emphasis here is on the company as opposed to a static and likely irrelevant scoring key.

Because of the unique nature of the scoring for the Company Service Perspective section, it is not included with the overall percent match for the CSP. Instead, it is used as a means by which a hiring manager or other corporate professional may make decisions concerning training needs, selection appropriateness, and performance expectations.

Statistical analysis is also a different matter for this section of the CSP as no norms are generated before the utilization of the product. What is measured is how similar a respondent's answers are to those selected by the company. Since this score does not contribute to the overall match percentage, reliabilities are not needed.

In developing the assessment, we created several question sets that address the unique aspects of five customer service environments including Health Care, Financial Services, Hospitality, Retail Services, and a General Customer Service category.

These five Company Service Perspective question sets were determined to fit the structure of many employers

in the United States. By reviewing our original list of Company Service Perspective questions and conferring with experts in the Customer Service industry, each question set was redesigned and augmented to possess greater utility for each environment.

This process began with approximately 150 questions related to serving the customer. A panel of employment professionals selected the 74 questions most descriptive of serving the customer.

Representatives from several industries were asked to rate each of the 74 questions on two issues of quality:

1. In your industry, which answer would be best, Yes or No?
2. On a Likert scale of one (not relevant) to three (very relevant), how does each question relate to your industry?

This information provided us with the best data for identifying how to tailor our pool of questions for each industry. The responses to these questions fell into five distinct groups with 50 questions for each industry:

- Health Care
- Financial Services
- Hospitality
- Retail Services
- General Customer Service

Each of the five question sets addressed aspects of customer service important to each industry.

The ability to specify one's company perspective will help to identify candidates who share many, if not all, of one's vision of good customer service ideals. This company-specific answer key feature helps tailor the assessment to one's needs. It also precludes the typical analyses one would conduct on a more traditional scale. No matching is performed; therefore, issues of reliability and norming are not necessary.

Chapter 5: Current Developments in the CSP II

As with all our products, ongoing validation of the Customer Service Profile II is a necessary element in our assessment development plans. By assessing our items, the way they interact with one another as evidenced by ongoing statistical analyses of real-world data, and creating a more utilitarian product, the demands of quality control and customer service are addressed.

To fulfill these demands, we focused on the items used for each of the six Behavioral Traits scales of the Customer Service Profile, and while the reliability of each scale was still satisfactory, it was decided that additional items per scale could only improve the quality of the product. Additional items also helped to enhance construct validity, as the increased number of items per scale can help to ensure the trait being measured is better defined.

For each scale, we created a number of items, themed for each construct by our team of professional psychologists. Our goal was to achieve 15 excellent items per scale, knowing that adding new items per scale in our initial research version would help to ensure finding the best while excluding those items that did not satisfy reliability after statistical analyses were completed.

We administered the new set of questions (all the original Customer Service Perspective Behavioral Characteristics items plus our new items) to a sample

of 4,633 adults in a corporate setting in July 2008. These adults included men and women in various sales, administration, technology, and operations positions. An analysis of reliability for each was generated, which identified the best 15 items per scale.

That list of items was then analyzed for reliability as the final version of the Customer Service Perspective, now called the Customer Service Profile II.

The current CSP II was completed by a sample of 55,300 participants between April 1, 2017, and July 31, 2020. Their results are shown in Tables 5.1 and 5.2.

Table 5.1

CSP Current Norm Sample Distribution (N=55,300)

Gender	Frequency	Percent
Male	20716	37.5
Female	30195	54.6
Missing	4389	7.9
Ethnicity	**Frequency**	**Percent**
Asian or Pacific Islander	1152	2.1
Black (not of Hispanic origin)	6826	12.3
Hispanic	13808	25.0
White (not of Hispanic origin)	27389	49.5
Missing	6125	11.1
Age Range	**Frequency**	**Percent**
12 - 15	170	0.3
16 - 25	22681	41.0
26 - 35	13358	24.2
36 - 45	6824	12.3
46 - 55	4962	9.0
56 - 65	2331	4.2
66+	391	0.7
Missing	4583	8.3

Table 5.2

CSP Behavioral Traits descriptive statistics

N=55,300	Trust	Tact	Empathy	Conformity	Focus	Flexibility
Mean	11.99	10.27	12.01	11.94	13.38	12.00
Std. Error of Mean	.01	.01	.01	.01	.01	.01
Std. Deviation	2.81	2.57	1.84	2.25	1.82	1.78
Minimum	.00	.00	1	.00	1	0
Maximum	15	15	15	15	15	15

As with the earlier version of the CSP, internal correlations of each scale help to identify how unique each scale may be compared to one another within the CSP II. Table 5.3 lists these correlations. All correlations are statistically significant. The scales that are expected to be related in a positive manner, where higher results on one are related to higher results on the other, are found for Trust and Tact, among others. The scales whose results are expected to oppose each other, such as Flexibility and Conformity, are negatively correlated, as expected. These results suggest a high degree of construct validity among the CSP II scales.

Table 5.3

CSP Behavioral Traits inter-scale correlation statistics

N=55,300	Trust	Tact	Empathy	Conformity	Focus	Flexibility
Trust	1.000	.372	.261	.430	.379	.040
Tact		1.000	.212	.365	.233	-.056
Empathy			1.000	.129	.141	.178
Conformity				1.000	.395	-.153
Focus					1.000	.161
Flexibility						1.000

All correlations are significant at the 0.01 level (2-tailed).

Table 5.4 shows the current alpha reliability coefficients. As can be seen, all are well above the recommended guideline of .70. The average reliability coefficient for the CSP is .80.

Table 5.4

CSP Behavioral Traits reliability coefficients (N=17,215)

Scale	Reliability	Scale	Reliability	Scale	Reliability
Trust	.78	Empathy	.82	Focus	.80
Tact	.79	Conformity	.80	Flexibility	.79
Vocabulary	.78	Numerical	.83		

Chapter 6: Summary and Conclusions

With regard to the internal validity of the CSP, an analysis of the relationships among the personality scales indicated a pattern of converging and diverging relationships that are clearly consistent with expectations based on the constructs being measured. The statistics clearly demonstrate positive correlations between related scales and negative correlations between two opposite scales.

Overall, results of our analyses of validity consistently indicate the CSP II is a valid measure of what it was designed and intended to measure, namely, six core Behavioral Traits related to customer service performance and two work-related Numerical and Vocabulary Proficiencies.

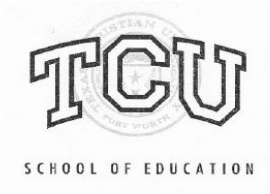
This report summarized the results of several validation projects that examined the relationship between the CSP scales and a diverse range of criteria. The results of these projects provided consistent and substantial support for the validity of the CSP II. Ratios of selection utilizing CSP overall match percentages when compared to measures of successful service performance were consistently accurate. These studies support the criterion validity of the CSP when job performance is the standard.

The reliability of the CSP II scales has been established through a variety of analyses, all of which demonstrate

the consistency the each of the six Behavioral Traits scales and the two Proficiencies scales.

In conclusion, statistical analysis of the CSP II clearly indicates its scales provide an accurate and reliable measurement of six core customer service related Behavioral Traits and two primary work-related Proficiencies. In addition, the Company Service Perspective (identifying alignment with your organizational culture) has been constructed to predict a number of important training-related issues and may act as a reliable guide for employee development.

Appendix A: Independent Review



May 30, 2000
To Dr. David W. Pearson:

It has given me great pleasure to review the questions and logic employed by you in the development of the Word Recognition testing instrument. I believe the program will provide all employers with an excellent tool for determining the verbal skills of job candidates and current employees.

The programs are cleverly designed with a sensible selection of "easy" words to allow even a less-educated person to feel a degree of success. At the same time, sufficient words fall into a moderate to difficult range. Overall, a very adequate basis is provided for employers to make appropriate selections based on word knowledge.

Throughout, the word meanings are abundantly clear and the correct answers distinct from the alternates without any unnecessary confusion or ambiguity. This holds equally true where the questions relate to use of correct synonyms or antonyms.

The usage of the words examined is all in current employ and the correct meaning is clear in every case. I found no bias or attempt to unfairly screen in any stage of the program.

As someone who regularly examines and employs contemporary testing related to literary education and word knowledge, I was delighted to find your Word Recognition testing instrument to be such a competent, thorough and fair evaluation tool.

Sincerely,

Luther B. Clegg
Professor & Chair
Department of Curriculum and Instruction
Texas Christian University

Texas Christian University • TCU Box 297900 • Fort Worth, Texas 76129

Appendix B: The Efficacy of Performance Modeling

As briefly mentioned in the Validity section, the assessment is a good example of how traditional thinking regarding test scores cannot apply when endeavoring to identify successful performers. Often, analysis of scores and criteria generates patterns for top performers that *do not* equate to consistently high scores on all scales of an assessment. In simpler terms, *higher is not necessarily better* when it comes to pattern matching systems. Nonetheless, the same would also apply for consistently low scores.

In this report, the details that support this actuality will be reviewed, which should help to guide the user when making future decisions concerning the patterns generated for his or her successful employees. While complex statistics may be referenced, summaries and implications will be stated that aid all users of the CSP II, no matter one's proficiency with statistical analysis or other assessment issues.

Performance Models are developed by comparing two forms of data. These data include the assessment scores acquired by employees in the organization for whom the model is being developed as well as the performance ratings of the same employees. Performance ratings are referred to as the criterion to which test scores are compared. When employees are defined into broad groups by their performance ratings, their assessment scores may be compared to identify

score differences in performance groups. These score differences define what may be considered effective identifiers that differentiate performance. The process identifies which of the scales individually differentiates between top and bottom performers.

For example, on a single scale in a test, let us say that performance groups (top and bottom performers) appear to have a tendency to score in different ways on the scale. Possibly, the top performers tend to score between six and nine on the ten-point scale while bottom performers tend to score from three to seven. This difference in score tendencies is a good sign that we may predict the potential performance of a new respondent (say for instance, a job candidate). The candidate is rated as to how similar his or her score on the scale may be to the tendency of scores from top performers (the model to which one is matched).

Keeping this in mind, it may appear to some that, like so many assessments we have taken in the past, a higher score is better. However, what if the top performers tend to score lower on the scale than bottom performers? That would define what is referred to as an inverse relationship or one in which higher performance is predicted by lower scores; in this case, lower is better.

In either the higher-is-better case or the lower-is-better case, a simple correlation statistic would be serviceable for identifying the relationship between scores and performance ratings (the criterion). The correlation would yield a highly positive or negative correlation

coefficient (i.e., maybe one that lies between + or – .3 to .9 or more). For the layman, such a correlation suggests that some relationship exists between the two factors (the bigger the numeral, the stronger the relationship), while the + or – suggests whether the relationship is a higher-is-better or lower-is-better connection.

However, what if the scores of top performers tend to lie in the five to seven area and the bottom performers tend to obtain scores of seven to ten? Or, even more complicated, the bottom performers tend to score across the whole spectrum of scores, say two to nine? Correlation tests would suggest no relationship exists at all! By taking a good look at the score distributions, the top performers can clearly be identified and differentiated from the bottom performers with only a small degree of false-positive effects. This effect, by the way, is one in which a respondent is identified as a potential top performer, when in fact his or her score is not reflective of such a potential; he or she is actually a bottom performer.

In such a case, correlation statistics cannot identify the relationship because they focus only on possible linear relationships between scores and criteria. Linear refers to the higher-is-better and lower-is-better situations. On a graph, such relationships are drawn as lines that slope across the graph and are thus referred to as linear. However, the relationship identified in the previous paragraph is not linear. Therefore, to identify such a relationship with statistical tests (the job of the test developer), we must utilize a test of correlation

that does not compare the scores and performance ratings but compares the job match percentage and performance ratings. This comparison is linear but is not adversely affected by Performance Models that lie in the center of the scale. The job match percentage has already altered the scoring of the individual on all scales in the test by tabulating how closely to the pattern he or she may be. By utilizing a job match percentage, the statistician may compare the data that help to demonstrate the validity of the assessment.

Appendix C: A Commentary on Validity Issues

Dr. David Pearson

An attempt at a more definitive discussion may be in order regarding a number of terms including validity that some may be prone to use rather loosely and/or incorrectly.

To assist in this endeavor, the wise words of Dr. Robert Guion, recognized as one of the leading authorities–if not *the* leading authority and author of the *Principles for the Validation and Use of Personnel Selection Procedures*–provide some insight.

Thus, the following rather extensive use of Dr. Guion's remarks from his presidential address entitled *Open A New Window; Validities and Values in Psychological Measurement* provide a rare, unique, and highly professional view of industrial psychology as it impacts the corporate world and its most important asset–people!

"Definitions of Validity"

Validity, in the history of testing, has been a confused concept, although the basic ideas have been present from the beginning. Criterion-related, content, and construct validities were all implicit when Galton said, "one of the most important objects of measurement...is to obtain a general knowledge of

the capacities of man by sinking shafts, as it were, at a few critical points. In order to ascertain the best points for the purpose, the sets of measurements should be compared with an independent estimate of the man's powers. We may thus learn which of the measures are most instructive" (DuBois, 1970).

Criterion-related Validity

Criterion-related validity is the extent to which scores on one variable, usually a predictor, may be used to infer performance on a different and operationally independent variable called a criterion. For convenience we often speak of criterion-related validities in terms of correlation coefficients, but the statistic has nothing to do with the definition.

Construct Validity

When we speak of understanding, we are talking about construct validity, the degree to which scores may be used to infer how well a stated hypothetical construct describes individual differences among the people tested. Construct validity is not expressible in such simple terms as validity coefficients. It is a judgment based on many kinds of information and procedures followed in developing the test and there results of experiments testing specific implications of the construct The data used to judge the construct validity of a measure may also help validate the construct; as data accumulate, ideas about the construct may be modified. Describing a construct is more than merely naming it.

What has or can be said about validity? First, for any measure there are many types of validity, not just one. Second, it is erroneous shorthand to speak of the validities of the measures themselves when what we are really referring to are the validities of inferences from the measures.

When we speak of criterion-related validity, we refer to the use of test scores to infer criterion levels. When we speak of construct validity, we refer to the use of test scores to infer degrees to which a particular construct describes the persons, organizations, or objects measured. When we speak of content validity, we refer to the use of test scores to infer levels of achievement the persons, organizations, or objects would exhibit in the total domain. These are different facets, or aspects, of validity and all three kinds of inferences should be valid for most tests.

However, the kind of validity statement we seek in any given measurement situation depends upon the kinds of inferences we wish to make. This is fundamentally a value judgment. Third, in industrial psychology, the most frequently valued inference is the inference about future performance on a valued criterion. The valued criterion is usually (although not necessarily) either a sample of a performance domain or a performance construct that is identifiably different from the constructs measured by predictors. That is, in employment testing, the validity of major interest is the validity of the hypothesis of a predictive relationship

between test scores and job performance. Fourth, the validities of criteria need to be investigated.

Finally, I suggest that an employment test may provide a basis for inferences that have criterion-related validity, or construct validity, or content validity, or all of these, and still not be job-related.

Job-Relatedness

I am not willing to equate "validity" and "job-relatedness." Criterion-related validity is evidence of job-relatedness only if the criterion measure is a valid measure of overall job performance, an element or sample of performance, or a construct related to job behavior.

Construct validity is evidence of job-relatedness only if the construct is related to the job. To defend an employment test on the basis of construct validity, one must first argue from job or need analysis that a particular construct is related to job behavior; then he must show that the test has acceptable validity for measuring that construct. What he has at this point is the hypothesis that the construct, as operationally defined by the predictor, can be used to infer levels of valued job behavior. This is a hypothesis of criterion-related validity, and it fairly begs to be tested–"where technically feasible."

If it is clearly feasible to do the criterion-related study, it should be done. Where it is clearly not feasible to do the study, the defense of the predictor can rest on a combination of its construct validity and the rational

justification for the inclusion of the construct in the predictive hypothesis.

From just a logical or rational viewpoint, one must be able to identify the specific task or tasks one is expected to perform in order to meet satisfactory performance measures.

As a first step and in keeping with construct validity research reports, a Criterion-related Study was accomplished according to the requirements under the *Uniform Guidelines on Employee Selection Procedures* Section 14 B, entitled *Technical Standards for Criterion-Related Validity Studies*, including the job analysis requirements under (2) of the same section in a Concurrent Method.

Thus, the first major step for conducting construct validity studies was performed as required under Section 14 D. That is, job relatedness was demonstrated for the trait names for those job groups where it was technically feasible to do so–a first, but crucial, step in a needed series of steps.

Now, this first criterion-related (concurrent method) study established that there was a job relationship which "...should (did) show by empirical evidence that the selection procedure was (is) validly related to the construct/trait(s) and that the construct/trait(s) were (are) validly related to the performance of critical or important work behavior(s)." Further, "the relationship

between the construct/trait(s) as measured by the selection procedure and the related work behavior(s), as determined from the job analyses, should be supported by empirical evidence from one or more criterion-related (concurrent) studies involving the job or jobs in question." Section 14 D. (3).

This was done as the first criterion-related study and job analysis for the construct validity of the Customer Service Profile and established that there was indeed a relationship between a number of the traits. In other words, job relatedness of the questions and traits utilized in the CSP were found to be significantly the same as some of those traits and questions utilized in the validation study which was approved as valid for a major corporation.

The original validation research and report was presented to the EEOC as the result of a class action discrimination complaint filed September 19, 1975 against The Adolph Coors Company with respect to their purported violation of 703(a) of Title VII.

The subsequent Settlement Agreement dated April 22, 1977 is on file in the Federal Courthouse of Denver, Colorado and is annotated as follows: Settlement Agreement in the United States District Court for the District of Colorado, Civil Action No. 75-W-992.

Anyone with a technical and statistical background and a strong “need to know” can acquire all or part of the actual validity study that was presented and subsequently accepted by the referenced Federal

Court by contacting the courthouse archives in Denver, Colorado.

The raw data, such as performance measurement forms and test answer sheets, upon which the validation study was conducted, remain with the client as proprietary and confidential data, rendering replication of this study impossible. However, this study–unlike other validation studies–has been scrutinized, analyzed, and verified by a Chief Federal Court Judge and his retained expert advisors as meeting the Federal Uniform Guidelines for the development of testing instruments.

Dr. Pearson contributed to the development of the Customer Service Profile and was integral to its success. He was a psychologist with expertise in Research Methodology, Statistics, and Testing with emphasis upon the Effects of Stress on Performance, Attitude Change, Neurophysiology, Learning Theory, Behavior Modification, and Personality Assessment. Dr. Pearson has been published in the journals of Aerospace Medicine, Perceptual Motor Skills, Aviation Medicine, Psychophysiology, and Genetic Psychology. He has also been a contributor for Newsweek, the Wall Street Journal, and The Personnel Administrator. His consulting firm provided private "hands-on" consulting to major businesses, federal and state agencies, and attorneys across the United States involving employment practices.

He was one of the original authors of the EEOC's and OFCCP's guidelines on how to prepare acceptable Affirmative Action Plans (AAPs) and contributing editor to the Federal Uniform Guidelines on Employee Selection Procedures, 29 CFR 1607 (1979). In 1973, Dr. Pearson was an invited participant of the EEOC's Coordinating Council in Washington, D.C. with respect to the proposed guidelines on testing. The input resulted in the Federal Uniform Guidelines on Employee Selection Procedures. Dr. Pearson's comments can be found on pages 161–190 of the record of those hearings–November 15, 1973.

References

Cochran, W.G. and Cox, G. (1957). *Experimental designs*. New York: Wiley.

DuBois, P.H. (1970). *A history of psychological testing*. Boston: Allyn & Bacon, Inc.

Edwards, A. (1960). *Experimental design in psychological research* (rev. ed.). New York: Holt.

Equal Employment Opportunity Commission. (1978). The Office of Personnel Management, U.S. Department of Justice and U. S. Department of Labor (1979). *Uniform Guidelines on Employee Selection Procedures*.

Federer, W.T. (1955). *Experimental design*. New York: Macmillan.

Guenther, W.C. (1964). *Analysis of variance*. Englewood Cliffs, New Jersey: Prentice-Hall.

Guion, R.M. (n.d.). Excerpts from the presidential address entitled "*Open A New Window; Validities and Values in Psychological Measurement*" at installation as President of APA's Division 14.

Winer, B.J. (1962). *Statistical principles in experimental design.* New York: McGraw- Hill.