

# 2013

National Association of Certified  
Child Forensic Interviewers (NACCFI)  
P.O. Box 7146  
Fredericksburg, VA 22404

## The NACCFI Competency Examination Item & Distractor Analysis The Final Report



### The Competency

This study was conducted based on the recommended best practices of the National Commission for Certifying Agencies (NCCA). The NCCA was established in 1977 in cooperation with the federal government. The NCCA standards require that credentialing programs seeking accreditation develop their competency examinations in accordance with the universally accepted *Standards for Educational and Psychological Testing*, developed jointly by the: American Education Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurements in Education (NCME). The content validity and reliability of the competency examinations used in credentialing is directly linked to the successful completion of the practice job analysis and the item and distractor quantitative analysis.

# Examination Task Force (CETF)

This report represents the hard and dedicated work of the 2009 Competency Examination Task Force (CETF). The CETF first convened on June 5, 2009 in Richmond Virginia for a three day workgroup in which to construct the first certification competency examination for child forensic interviewers. The CETF was directly responsible for linking the exam specification and blueprint with the findings of the 2008 Practice Job Analysis. The CETF was also responsible for developing the initial item bank used for the current NACCFI certification competency examination. This report represents the outcome of their diligent work.

Hector M. Campos MSW, LCSW  
CETF Coordinator  
Fredericksburg, Virginia

Kelly Bober BA, DCFI  
CAC Executive Director  
Winchester, Virginia

Ann D. Goodall, LBSW, ACFI  
Social Work Specialist  
Child Forensic Interviewer  
Topeka, Kansas

Leslie Vassilaros MS, LSW, DCFI  
CAC Executive Director  
Wheeling, West Virginia

Christina Copland M.S. Ed.  
Child Forensic Interviewer  
Maryville, Tennessee

Lisa Mitton MSW, LCSW  
Child Forensic Interviewer  
Twin Falls, Idaho

Daryl B. Leach MSW, LCSW-C, DCFI  
Child Forensic Interviewer  
Rockville, Maryland

Michele Thames, DCFI  
Child Forensic Interviewer  
Norfolk, Virginia

Johanna Hager LCPC, DCFI  
CAC Executive Director  
Morrison, Illinois

Paul D. Naylor, PhD  
Psychometric Consultant  
Durham, North Carolina

# Table of Contents

## Report Headings:

The Competency Examination Task Force	Page 2
The Executive Summary	Page 4
The Evidence Based Process	Page 5
Statistical Measures Used	Page 6
The Results	Page 6
The Exam Form	Page 8
Item Difficulty	Page 9
Item Discrimination	Page 10
The Distractor Analysis	Page 11
The Examinee Sample	Page 12
Other Important Findings	Page 13
The Standard Setting	Page 14
In Conclusion	Page 14
References	Page 30

## Tables:

Table I: Table of Exam Specifications	Page 7
Table II: Exam Form Statistics	Page 8
Table III: Item Level Statistics	Page 10
Table IV: Item Difficulty by Grouping	Page 10
Table V: Item Discrimination by Grouping	Page 11
Table VI: Examinee's Sample Demographics	Page 12
Table VII: High Scoring Examinees Groupings	Page 13
Table VIII: Low Scoring Examinees Groupings	Page 13

## Appendix:

Appendix A: Statistical Analysis Worksheet Formulas	Page 16
Appendix B: The Examination Blueprint	Page 17
Appendix C: Examinees Scores Ranked & Ordered	Page 18
Appendix D: Item Difficulty Vales Ranked & Ordered	Page 21
Appendix E: Item Discrimination Values Ranked & Ordered	Page 25
Appendix F: Randomized Distractor Frequencies and Mean Scores	Page 27

## The Executive Summary

This report presents the findings of a detailed item analysis field study conducted by the National Association of Certified Child Forensic Interviewers (NACCFI). According to the *Standards for Educational and Psychological Testing*, the primary purpose of professional credentialing is to identify candidates who possess the core knowledge competencies and moral character believed necessary for safe and effective practice within a given profession. It further states that for this purpose, “tests used in credentialing are intended to provide the public with a dependable mechanism for identifying practitioners who have met these particular standards”. This study will examine the validity and reliability coefficients of the NACCFI examination form by measuring the psychometric internal consistency of each item in the form. The purpose for conducting this study is to determine if the current competency examination form used to credential child forensic interviewers, credibly measures the knowledge competencies it was designed to measure. This report will provide test level summary statistics based on the raw number correct scores for the total score of the 100 items in the exam form. The outcome competency data used to conduct this study was collected from 100 currently practicing child forensic interviewers who submitted a credentialing application and were found qualified to sit for the current child forensic interviewer certification competency examination, based on the current multi-tiered credentialing standards.

## Credentialing Core Requisites

There are three primary core requisites considered essential for all professional credentialing programs of excellence. They include; fidelity to practice, evidence of practice competency and evidence of the applicant having good moral character. Fidelity to practice requires the applicant to provide evidence of actual practice experience in the work of the profession and a commitment to participate in practice related continuing education training as a condition to maintain or renew their credential. Practice competency requires the applicant to show evidence of having the entry level core knowledge competencies and skill sets believed necessary to do the work of the profession safely and effectively. Practice competency is typically evidenced by the applicant demonstrating actual practice experience, professional training related to the practice and the application of a performance or competency examination requisite design to credibly measure the applicant’s knowledge competency. Good moral character requisites are designed to provide the consumers of the profession and the general public some reasonable assurances that the applicant is of good moral and ethical character, particularly when their practice involves working with children. Moral character is typically measured by a requisite for character endorsements from the applicant’s professional peers and supervisors, a criminal history background check to identify any past criminal behaviors, and a commitment by the applicant to abide by a code of professional principles, values and code of ethical conduct. This study will only address

one of the core credentialing requisites, the validity of the certification competency examination used to credential child forensic interviewers.

### The Evidenced Centered Design

The Evidence-Centered Design (ECD) is a construct for the development of evidenced based examination instruments used in credentialing. This design helps to clarify the rational and intended purpose of the examination and provides the psychometric empirical evidence needed to make appropriate conclusions about the instruments validity. Validity refers to the effectiveness and appropriateness of the exam instrument to be used for its intended purpose. Validity is not an inherent property of the examination instrument rather a continuing process of assessment from multiple sources. In order to credibly defend the validity of any competency examination instrument, the link between what is believed to be sufficient knowledge for entry level safe and effective practice must first be established using an evidenced centered design. This link can only be established by the successful completion of a comprehensive practice job analysis. (Mislevy, Almond, & Lukas, 2003, p. 20). (Lissits & Samuelsons) page 477).

### The Practice Job Analysis

The completion of well constructed practice job analysis is the first step in a series of recommended best practices for any organization, profession or discipline seeking to establish an accredited voluntary or statutory credentialing program. The outcome product of a practice job analysis is a detailed list of observable tasks, functions and core knowledge competencies believed necessary for safe and effective practice within a given profession. This data when properly analyzed serves as the empirical evidence for the construct of competency examination instruments, training programs and academic core curriculums associated with the profession.

The content validity of the competency examinations used for credentialing is directly linked to the successful completion of the practice job analysis. (Mehrens, 1995; Tannenbaum, 1999; Raymond, 2001; 2002; Raymond & Neustel, 2006). Between 1996 and 2009, over 25 cases were heard in US District Courts and the US Supreme Court concerning the content validity of competency examinations and their impact on employability. In every one of these cases, the practice job analysis was upheld as the essential element needed to establish content validity. The practice job analysis survey is designed to gather empirical data related to the performance of a job or practice directly from the front line practitioners who are actually doing the work of the profession. The survey uses a 3 part likert type scale designed to identify the task and functions believed most critical to effective and safe practice. The scales ask the participants to identify how important a task is, how frequently it is performed and how much it is needed in order the practice safely and effectively. The sum of these scales is then calculated to identify those functions that are believed most critical to effective and safe

practice. The first practice job analysis for child forensic interviewers commenced with the appointment of the Practice Analysis Task Force (PATF) in 2008. In the 2008 practice analysis survey over 305 currently practicing child forensic interviewers participated in a validation survey designed to identify the critical tasks and functions associated with safe and effective practice in child forensic interviewing. In a second follow up practice analysis that commenced in 2011, over 1000 currently practicing child forensic interviewers validated the findings of the 2008 study. (2008 Practice Job Analysis)

### The Item Analysis Results

The overall findings of this study indicates that the NACCFI certification exam form is fully stabilized and closely aligned with the targets identified in the table of specification and the examination blueprint. The findings further signal strong fidelity between the content of the exam items and the current practices of child forensic interviewers across a variety of geographical and practice settings. These findings are evidenced by a reliability coefficient value of (KR-20=.83). The (Kuder Richardson Reliability Index, 1937) commonly referred to as the (KR-20) is the most widely used index to measure internal consistency and reliability for high stakes examinations. KR-20 scores theoretically range in value from 0.0 which means no reliability to 1.0 which indicates perfect reliability. Higher reliability scores indicate that the items tend to correlate well with each other. A low KR-20 score means that the items tended to be unrelated to each other and/or the actual knowledge competencies of the examinees. KR-20=.80 values and higher are deemed psychometrically sound and reliable. Lower scores may also be acceptable, however they may also signal that the overall scores have not established a pattern of consistency and stability which may or may not be related to internal errors in the items. A (KR-20=.83) value suggest minimal internal errors in the items as the overall scores have established a positive pattern of consistency, stability and reliability. Typically the KR-20 scores are higher for exam forms with larger samples of 200 or more. For the current sample of 100 examinees, a KR-20=.83 is a reliable indicator of the exams form validity. This validity claim is further validated by the point biserial co-efficient for each of the 100 items and the fidelity of the item difficulty (P) coefficients with the targets identified in the examination table of specification and blueprint. The point biserial correlation (rpb) is an index that measures the power of each individual item to discriminate examinees with higher and lower Knowledge Skills and Abilities (KSA). As the findings indicate, all of the 100 items in the current NACCFI examination form had a point biserial coefficient greater than .10 (rpb>.10).

### The National Accreditation Standards

This study was conducted in compliance with the recommended standards of the National Commission for Certifying Agencies (NCCA). The NCCA is responsible for promoting and enforcing regulatory standards of excellence for voluntary and statutory credentialing programs. The NCCA was established in 1977 in cooperation with the federal government. The NCCA standards requires credentialing programs seeking accreditation to publish a description of the assessment instruments used to make certification decisions, as well as the research methods used to ensure that the assessment instruments are valid, in accordance with the universally accepted *Standards for Educational and Psychological Testing*, developed jointly by the American Education Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurements in Education (NCME). According to the NCCA, it is possible to collect sufficient evidence to ensure that the examination results are appropriate for their intended purpose and therefore assert claims of validity.

### The Competency Examination Task Force (CETF)

On June 5, 2009 a ten member Competency Examination Task Force (CETF) convened in Richmond, Virginia for a three day work group tasked to construct the first certification competency examination form for child forensic interviewers. The CETF members were selected from a list of applicants that responded to an open call for nominations. Each member was selected primarily due to their Level of direct practice experience in child forensic interviewing and from diverse practice settings and geographic locations. The CETF members were mentored and trained in the science of developing examination blueprints and item writing by Dr. Paul Naylor, a recognized expert in the field of occupational analysis and psychometric testing. The CETF was tasked to use the literature review, the role delineation study and the validation survey data to develop evidence informed content categories for the examination blueprint and the table of exam specifications. The CETF was also instructed on the principles of examination item development, and the reliability and content validity indexes required to validate the items and the exam form. The number of items that were assigned to each content category was decided by the weight of their overall critical ranking and the frequency of knowledge competencies assigned to each critical task. The content outline resulted in the examination blueprint for the current examination form. The examination blueprint identifies the content headings and subheadings for each competency being measured and the percentage of items relating to each of the competencies based on their total overall weight value and the skills level targeted for that item. The table of exam specifications outlines the descriptive targets the exam form was designed to measure. See **(Appendix B)** the table of exam specification and **(Appendix C)** the examination blueprint.

### The Statistical Measures Used

This item analysis study was conducted using Classical Test Theory (CTT) measurements. The CTT model is the easiest and most widely used model for the analysis of competency examinations particularly when working with smaller samples. The Statistical Analysis Worksheet (SAW) is an excel worksheet designed by NACCFI to provide detailed item and distractor level data using the CTT measurement model. The SAW was also designed to use regression analysis to examine each item and distractor choice in reference with other variables associated with the examinee. This also allows us to identify variables that significantly correlate with higher and lower competency outcome scores. The output data is provided in two separate files; a detailed file of item and distractor analysis statistics, and a file of examinee scores and exam Level form statistics. **(Appendix A)** lists the statistical formulas used to conduct this item and distractor analysis.

### The Examination Form

As indicated in **(Table II)**, 100 examinees qualified to sit for the NACCFI competency examination. The exam form contained 100 items with one stem, one key, three distractors and only one possible correct answer choice. The cut score for this exam form was initially set at 60% which was an actual score not a weighted score. The minimal score recorded was 64, which mean that the lowest scoring examinee answered 64 of the 100 items correctly. The maximum score was 96, which indicates that the highest scoring examinee answered 96 of the 100 items correctly. The mean score from all 100 examinees was 84 (averaged). The Standard Deviation (SD) is 7.74. The SD is a measure of the dispersion of total scores. The SD indicates how "spread out" the scores were from the mean. The standard deviation is also a reliability measure which is closely related to the Standard Error Measurement (SEM), which is inversely related to reliability. The SEM for this exam form is 3.19. The Standard Error Measurement (SEM) is a reliability coefficient of internal errors expected in all exams forms from a variety of sources to include internal and external factors. The SEM is inherent in all exams and is associated with circumstances that cannot be controlled during the examination process. The error can occur in the plus or minus of the observed score. For this exam it appears that the examinees true scores will normally fall within plus or minus 3 intervals of their observed scores. If the SEM is small, this means that there is a more accurate estimate of the true score. If the SEM is large, there is not an accurate estimate. The expected SEM for all exams is 2.0. The usefulness of the SEM score is that it provides an index for exam score accuracy. The SEM is also critical as it is factored in the standard setting process used to determine the exams cut score.



**Table II: Exam Form Statistics**

Examinees	100
Items	100
Mean Score	83.85
Standard Deviation (SD)	7.74
Variance	59.85
Minimal Score	64
Maximum Score	96
KR-20	0.83
SEM	3.19

The Examinee Sample

A sample of 100 actual examinee scores was used to conduct this item and distractor analysis. The first examinee to sit for the competency examination was scheduled in January 2011. The last examinee completed the exam on July 2012. Each examinee in the sample successfully applied for and was qualified to sit for the certification competency examination based on the current NACCFI multi-tiered credentialing standards. This process ensures that the field sample was fully representative of the actual practitioners eligible and qualified to receive the NACCFI credential. No course of study or additional training in preparation for the exam was offered. This decision ensures that the examination scores are an accurate measure of current practice competency and not our training or exam preparations program. The ACT Center Network was contracted to deliver and proctor the exam. Proctoring is a recommended best practice to ensure that the security of the exam and its content is not compromised. The ACT network has a current capacity of over 1,400 work stations available nationwide throughout the year. TesTrac an independent testing service was contracted to independently administer the exam and collect the raw outcome data. (Table VI) outline the examinee sample population.

**Table VI: Examinees Sample Demographics**

Gender	Reported Training		Completed Interviews		Practice Setting		Years in Practice	
82 Female	41-80	31	25-50	26	CAC	61	0-2	19
18 Male	81-120	22	51-100	13	CPS	22	2-5	43
	121-160	8	101-500	33	Police	11	5-10	25
	161-200	9	5001-1000	32	Courts	4	10-20	9
	201-400	6	1000 +	6	Private Practice	2	20 +	4
	401-800	21						
	801-1000	3						

### Item Difficulty the (*P*) Value

When conducting an item analysis, there are two values that we are interested in measuring at the item level. These values are item difficulty and item discrimination. In the CTT model, item difficulty is represented by the (*P*) value. The *P*-value simply measures the proportion of examinees that correctly answered an item with one correct alternative worth a single point. The item difficulty index ranges from 0 to 1.0 or 100. If only 50% percent of the examinees answered the item correctly, it is considered a difficult item, if 90 % of the examinees identified the correct answer, then the item is considered easy. The higher the *P* value, the easier the item. Item difficulty plays an important role in the ability of an item to discriminate between examinees who know the material, and those who do not. This measure is done by analyzing the items behavior empirically, and contrasting the findings with the identified targets in the table of specifications. For most credentialing programs, items with a  $P < .50$  are considered difficult items, as only 50% or less of the examinees are answering the item correctly. A value of  $P = 0.0$  means that none of the examinees answered the item correctly and the item may be too difficult or has no correct answer. The table of specifications for this exam form targeted 75 easier Level I items; with a  $P = .70 < 100$ , and 25 more difficult Level II, items; with  $P = .40 < .70$ . Level I items rely more on the simple recognition of learned facts and theories. Level II items require the application of reasoning and synthesis, with a genuine understanding of the concepts and some practice experience. The number of items for each Level was determined by whether the content area required simple identification of learned knowledge, or higher Level reasoning and application skills. How difficult or easy an item may be is a judgment call that has to made, while taking into account the content of the item, the overall mean score response and the targeted abilities of the examinees (Shultz, K.S., & Whitney, D.J. (2005). The actual targeted *P* values for this exam were established by the CETF based on the skill Levels assigned to each of the content areas in the examination blueprint. The examination blueprint identified 33 of the items in the form as Level II items, of these only 17 met the Level II designation. The current distribution of the difficulty items for the NACCFI examination form comes very close to what is desirable, covering practically the whole gamut of *P* values. As indicated, the mean *P* value for the all the items is .84. The present findings indicate a slightly greater proportion of easier items in contrast to the targets identified in the table of specification. 83 of the items fall under the Level I category in contrast to the 75 targeted in the table of specifications, only 17 items fall in the difficult Level II range, with no items identified as very difficult. The fact that the difficulty targets do not exactly correlate with the targets identified in the table of specification does not diminish the reliability or validity of the exam form, primarily because this exam form is targeted for entry Level practitioners with a minimum of 25 completed interviews. The actual examinee sample indicates a majority of the examinees 84% are identified as advanced and Diplomate Level practitioners. This indicates that this exam form is actually more difficult than anticipated for the entry Level practitioner. The process of establishing a balanced sample of

easy and more difficult items is common to all new examination programs. The issue is addressed by the development and empirical validation of new items and the construct of multiple exam forms with varying targeted ranges of item difficulty. **(Table III)** is a listing of all the item level statistics. **(Table IV)** identifies the items by their difficulty groupings. **(Appendix D)** shows the complete listing of all the items ranked and ordered by their (*P*) Value.

**Table III: Item Level Statistics**

Mean <i>p</i>	.84
Minimal <i>p</i>	.35
Maximum <i>p</i>	.99
Mean <i>Rpbis</i>	0.27
Minimal <i>Rpbis</i>	0.10
Maximum <i>Rpbis</i>	0.53
Item Count <i>p</i> = .70<.100	83
Item Count <i>p</i> = .40<.70	17

**Table IV: Item Difficulty by Groupings**

Very Difficult ( <i>P</i> < 0.3)	Moderately Difficult ( <i>P</i> = 0.31-0.50)	Difficult Items ( <i>P</i> = 0.51-0.70)	Moderately Easy ( <i>P</i> = 0.71-0.80)	Easy ( <i>P</i> > 0.80)
0 (0%)	7 (7%)	10 (10%)	19 (19%)	64 (64%)

#### Item Discrimination Coefficients

Perhaps the most important function of a competency examination is its ability to differentiate between those examinees that know the knowledge competencies identified as necessary for safe and effective practice, and those examinees that do not. Item discrimination is an index that measures the power of the item to discriminate between examinees with higher KSA's, and those with lower KSA's. Item discrimination correlations can be built into the examination by choosing items with higher or adequate discriminative values. The point biserial correlation ( $r_{pb}$ ) is an index that measures the power of the item to discriminate between examinees with higher KSA's and those with lower KSA's. Item discrimination is typically defined as the correlation between item scores and the total scores. An ( $r_{pb}$ ) of 0.0 indicates that there is no correlation or relationship between the item difficulty and the total scores of those who answered the item correctly. It also indicates that the items are unable to discriminate between higher and lower skilled examinees and/or that the item responses are essentially random with respect to the total scores. A negative or minus - ( $r_{pb}$ ) is problematic because there is

an inverse relationship in scores, indicating that lower scoring examinees identified the correct option more frequently than higher scoring examinees. This typically means that there is a key error, no correct answer, more than one correct answer, and/or the examinees are randomly guessing the correct option. A perfect ( $r_{pb}$ ) value of 1.0 indicates that everyone who answered the item correctly had a higher overall exam score than everyone who answered the item incorrectly. Choosing items with higher ( $r_{pb}$ ) maximizes the reliability of the examination form. The findings for this analysis indicate very positive discrimination values, evidenced by 79% of the items sharing a very positive or excellent discriminatory value and 21 % of the items indicate a positive discrimination value meaning none of the items discriminated lower than .10 or negatively. As indicated in **(Table III)** above, the mean ( $r_{pb}$ ) is .27, the lowest ( $r_{pb}$ ) is 0.10 and the highest ( $r_{pb}$ ) is .53. **(Table V)** below, is a grouping of the items discriminatory values. **(Appendix E)** shows a listing of the 100 items ranked and ordered according to their ( $r_{pb}$ ) value.

**Table V: Item Discrimination Values by Grouping**

Discriminate Poorly ( $r_{pb} < 0$ )	Positive Values ( $r_{pb} = .10 - .14$ )	Very Positive Values ( $r_{pb} = .15 - .35$ )	Excellent Values ( $r_{pb} > 0.35$ )
0 (0%)	21 (21%)	46 (46%)	33 (33%)

### The Distractor Analysis Data

When attempting to identify flawed items, it is very important to examine the distractors that are being identified as the correct answer by those examinees who answered the item incorrectly. An analysis of the distractors shows the mean scores of the examinees that chose the correct and incorrect options for each item. It also provides the frequency by which distractors are more inviting to those examinees that exhibit lower and higher mean scores. This information is vital particularly in determining whether the items are effectively discriminating between the higher scoring examinees, and the lower scoring examinees. The goal of conducting the distractor analysis is to ensure that more than one of the distractors presented was attractive, plausible or believable to those examinees that had not mastered the concepts being evaluated; especially if it is a commonly known and accepted practice principles that most entry Level practitioners should know. The items are also evaluated to ensure that all distractors provided some reasonable correlation to the competencies being assessed and the correct key. All of the items that asked for the MOST, LEAST or PREFERRED responses were capitalized to minimize confusion. The distractor analysis also helped identify items that had two possible correct answers, and two items that were keyed incorrectly. Correlations between examinee responses and their

training received can help identify gaps in training and/or difficult tasks, functions and concepts that require more training to master. By correlating the individual distractor choices for each examinee with other variables such as the types of training received, practice experience, practice settings and other demographics variables, we can isolate variables associated with higher and lower competency scores. **(Appendix F)** provides a listing of all the items, the frequency values and the means scores for the examinees that identified the correct key and the incorrect distractor for each item.

### Other Important Findings

This study also conducted a very preliminary investigation using co-relational statistics to determine if there were any variables that significantly correlate with higher or lower competency outcome scores. This data is critical in supporting the other prerequisites standards associated with receiving the NACCFI certification credential. This investigation was conducted by a preliminary examination of some examinee related variables associated with a sample of the highest scoring examinees and a sample of the lowest scoring examinees. The determination classification for higher and lower ability examinees was done by calculating the mean for the 25 highest, and 25 lowest outcome scores. Variables associated with the number of reported training hours received, the number of reported interviews completed and the reported practice settings were each of the examinee practiced were then correlated for each of the high and low scoring examinees in the sample. The preliminary findings identified a strong correlation between increased training hours reported and the higher competency scoring sample. The findings also found significantly higher competency scores for those examinees who participated in training associated with more than one interview protocol or model. All of the examinees representing the high scores participated in one or more of the RACTAC model training and one or more variants of the NICHD training protocol interview model. All of the examinees in the higher scoring sample work for a Child Advocacy Center (CAC) and all had completed a minimal of 160 hours of protocol specific training in more than one of the current nationally accepted interview protocol structures models. The findings did not identify any significant correlation between higher competency scores and academic degrees received or the number of interviews completed. It is important to consider that preliminary study is only designed to measure the practice knowledge competency. While some of these variables may not correlate significantly with the measures of knowledge competency, they may correlate with increased Levels of performance skills associated with conducting the child forensic interview. **(Table VII) & (Table VIII)** present sample demographics of the higher and lower scoring examinees **(Appendix C)** shows the examinees means scores ranked and ordered from higher to lower. The high and low 25 scores are highlighted in green.

**Table VII: Sample Demographics for High Scoring Examinees**

Mean Score	Gender	Reported Training Hours	Completed Interviews	Practice Setting	Years in Practice					
93.88	Female	25	161-200	9	101-500	10	CAC	25	0-2	1
	Male	0	401-800	16	501-1000	11	CPS	0	2-5	16
					1000 +	4	Police	0	5-10	8

**Table VIII: Sample Demographics for Low Scoring Examinees**

Mean Score	Gender	Reported Training Hours	Completed Interviews	Practice Setting	Years in Practice					
74.08	Female	23	41-80	24	25-50	18	CAC	16	0-2	18
	Male	2	81-120	1	51-100	5	CPS	7	2-5	5
					101-500	2	Police	2	5-10	2

The Standard Setting

Examinations used in credentialing are only expected to cover the knowledge competencies that must be present upon entry into a profession in order to practice safely and effectively. Nonetheless, it is this connection between sufficient knowledge for job performance and content knowledge covered by a credentialing exam that must exist in order to support the validity of the credentialing exam scores. (AERA, APA, & NCME, 1999).

The standard setting is a term used to describe the passing score or the cut score for the competency examination. The standard setting represents the performance setting, or the passing score required by an examinee as evidence that they have the sufficient knowledge competencies believed necessary to practice safely and effectively. For the purposes of accreditation, the NCCA standards require certification programs to set their cut score based on a criterion referenced approach. Based on the current data provided by this study, the passing rate or cut score for the NACCFI certification examination will reference the criterion provided from this study. To ensure that no examinee was unduly affected by any initial flaws on the exam form, the initial cut score or this exam was set at 60% and the recommended passing score was established at 75 %. These are actual scores not scaled scores. Based on the findings from this study the revised cut score for the NACCFI Competency examination form has been increased to the suggested 75%. This score is derived by the mean score for all examinees 83% minus the SEM .3 and a .5 adjustment related to the differences to the sample population. As indicated credentialing exam forms are designed to measure entry level competencies and the sample used for this study was represented by 83% advanced and diplomate examines. The actual mean score for 25% of the low scores is 75%. These absolute standard settings are based on empirical examinee data and represent a more reliable process than the initial normative settings established for this exam. Our goal is to construct an item bank with a minimal of 500 validated items prior to convening an Angoff Study to establish

the final cut scores for more than one exam form. In order to increase the item bank with reliable field tested items the exam form has been increased to 150 items, 50 of which are being field tested and will not count towards the examinees final score.

### In Conclusion

This report represents one piece of a much larger puzzle, the process of establishing the credibility and effectiveness of any credentialing program must include psychometric evidence that the examinations instruments being used are psychometrically sound and thus legally defensible. While the successful completion of a practice job analysis is a requisite to the construct of the exam form, the validity of the exam form in turn supports the reliability of the practice job analysis process. A valid examination form indicates that the current practices of child forensic interviewers were credibly lineated in the practice job analysis, and shares credible fidelity to the actual practices of the profession. It is equally important to note that while a professional credential signifies that the practitioner met the requisites believed necessary for competent practice, it is not a guarantee that the practitioners will practice ethically or competently. According to Schmitt, 1995, a credential is neither a guarantee of the public's protection nor of the practitioner's competency on the job. Clauser et al. attribute this distinction, in part to the difference between knowing what to do and actually doing it. (Clauser, Margolis, and Case (2007).

# Appendix A: Statistical Analysis Worksheet (SAW) Formulas

<i>facility</i>	<i>P+</i>		
	<i>P-</i>		
<i>difficulty</i>	<i>Examinees correct</i>		
	<i>Examinees incorrect</i>		
	<i>Mean All</i>	83.85	
	<b>STD DEV all</b>	<b>7.74</b>	
	<i>M+</i>		
	<i>M-</i>		
	<i>M-T</i>		
	<i>Sdev</i>		
	$(M+ - M-T)/Sdev$		
	$SqRoot(P+/P-)$		
	<b>Rpbis</b>		
	<i>n</i>		100
	<i>Xave</i>		83.85
	$X^2/n$		70.31
	$S^2$		59.85
	$(X-X^2/n)/S^2$		0.23
	$(1 - ( ))$		0.77
	<b>KR-21</b>		<b>0.78</b>
	<i>KR-20</i>		
	$k=n=100$		
	$\Sigma p*q$		0.14
	$1-(\Sigma p*q/\sigma^2)$		0.82
	<b>KR-20</b>		<b>0.83</b>
	$(SQRT(1-KR20))*STDdev$		
	$SQRT(1-KR20)$		0.41
	<b>SEM</b>		<b>3.19</b>
		<i>ave</i>	
	<i>IF-</i>	0.84	
	<i>IF+</i>	0.83	



## Appendix B: Table of Exam Specifications

Exam Description	Child Forensic Interviewer Entry Level Certification Competency Examination
Purpose of the Exam	Assessing the foundational core knowledge competencies believed necessary for effective and safe practice as an entry-Level child forensic interviewer.
Length of the Exam	100 items
Item Format	All items are multiple choices with one stem, a single key and three distractors.
Target Population	Entry Level child forensic interviewers who has conducted a minimal of 25 child forensic interviews.
Cognitive Skills Levels being Evaluated	Level I: Recognition of basic interview structures, protocols and recommended best practices as indicated in the literature and validated by the practice analysis.  Level II: The application of basic assessment, problem solving and reasoning requiring practice knowledge of recommended best practices for child forensic interviewers.
Number of Items Targeted Per Skill Level	Level I = 75 items Level II = 25 items
Targeted Difficulty Index for each Level	Level I: $P = .70 < 100$ Level II: $P = .40 < .70$
Time Allocated to Complete the Exam	Two hours and thirty minutes
Exam Administration, Review and Delivery	TesTrac LTD. Professional Examination Services
Exam Proctoring and Security	The ACT Testing Network

## Appendix C: The Examination Form Blueprint

Knowledge Content Areas:	Weight	Skill Level	# of Items
<b>01: Dynamics of Child Maltreatment:</b>	<b>14%</b>		
KSA 01: Dynamics of Intra Familial Child Sexual Abuse	0.03876	Level: I	6
KSA 25: Interview Barriers & Interventions	0.03777	Level: II	6
KSA 03: Child Prostitution & Human Trafficking	0.00551	Level: I	1
KSA 04: Pornography and Internet Exploitation	0.00551	Level: I	1
<b>02: Child Developmental Theory:</b>	<b>30%</b>		
KSA 10: Memory and Suggestibility	0.04382	Level: II	5
KSA 08: Comprehension & Recall	0.09338	Level: I	10
KSA 09: Language and Articulation	0.06357	Level: I	6
KSA 02: Developmental Disabilities	0.05208	Level: II	5
KSA 17: Cultural Diversity and Sensitivity	0.04698	Level: I	4
<b>03: Forensic Interview Structures, Protocols &amp; Aids:</b>	<b>30%</b>		
KSA 05: Interview Protocols	0.14583	Level: I	14
KSA 24: Interview Structures	0.04960	Level: I	4
KSA 23: Interview Aids and Props	0.02583	Level: II	2
KSA 11: Question Typology	0.10323	Level: I	10
<b>04: Criminal Child Abuse Investigations Procedures:</b>	<b>09%</b>		
KSA 20: The Multidisciplinary Team Approach:	0.08372	Level: I	4
KSA 12: Rules of Crime Scene Evidence	0.02708	Level: I	2
KSA 18: Corroborating Evidence	0.00753	Level: II	2
KSA 19: Alternative Hypothesis	0.00630	Level: II	1
<b>05: Court Procedures &amp; Case Law:</b>	<b>17%</b>		
KSA 21: Court Testimony Case Law	0.05783	Level: I	3
KSA 16: Child Maltreatment Case Law	0.01577	Level: I	3
KSA 13: Court Testimony Procedures	0.01733	Level: II	7
KSA 07: Child Victim Advocacy	0.01368	Level: I	1
KSA 22: Ethical Considerations	0.05717	Level: II	3
<b>Total Items:</b>			<b>100</b>

## Appendix D: Examinees Scores Ranked and Ordered

Examinees Scores			Item Statistics					Mean score correct	Mean score incorrect
Examinees	ID	Score	Item	<i>P</i>	<i>Rpbis</i>	Number Correct	Number Incorrect		
25	C00045	96	25	0.81	0.17	81	19	84.48	81.16
64	C00084	96	64	0.77	0.33	77	23	85.25	79.17
80	C00101	96	80	0.81	0.17	81	19	84.48	81.16
49	C00069	95	49	0.38	0.33	38	62	87.16	81.82
53	C00073	95	53	0.78	0.47	78	22	85.77	77.05
62	C00082	95	62	0.76	0.36	76	24	85.43	78.83
87	C00108	95	87	0.80	0.47	80	20	85.68	76.55
98	C00119	95	98	0.97	0.19	97	3	84.10	75.67
42	C00062	94	42	0.98	0.35	98	2	84.23	65.00
47	C00067	94	47	0.78	0.39	78	22	85.46	78.14
78	C00099	94	78	0.74	0.28	74	26	85.14	80.19
85	C00106	94	85	0.81	0.17	81	19	84.48	81.16
88	C00109	94	88	0.99	0.15	99	1	83.97	72.00
92	C00113	94	92	0.98	0.35	98	2	84.23	65.00
97	C00118	94	97	0.74	0.30	74	26	85.24	79.88
31	C00051	93	31	0.39	0.34	39	61	87.10	81.77
43	C00063	93	43	0.98	0.35	98	2	84.23	65.00
45	C00065	93	45	0.99	0.15	99	1	83.97	72.00
70	C00090	93	70	0.76	0.28	76	24	85.05	80.04
83	C00104	93	83	0.74	0.30	74	26	85.24	79.88
89	C00110	93	89	0.98	0.35	98	2	84.23	65.00
28	C00048	92	28	0.77	0.35	77	23	85.35	78.83
32	C00052	92	32	0.80	0.47	80	20	85.65	76.65
40	C00060	92	40	0.75	0.28	75	25	85.09	80.12
75	C00096	92	75	0.81	0.17	81	19	84.48	81.16
91	C00112	92	91	0.97	0.13	97	3	84.03	78.00
100	C00121	92	100	0.98	0.35	98	2	84.23	65.00
37	C00057	91	37	0.35	0.26	35	65	86.63	82.35
46	C00066	90	46	0.97	0.22	97	3	84.15	74.00
59	C00079	90	59	0.97	0.35	97	3	84.33	68.33
14	C00034	89	14	0.81	0.14	81	19	84.37	81.63
1	C00021	88	1	0.83	0.13	83	17	84.31	81.59
23	C00043	88	23	0.65	0.14	65	35	84.63	82.40
21	C00041	85	21	0.98	0.35	98	2	84.23	65.00
2	C00022	84	2	0.96	0.19	96	4	84.16	76.50
16	C00036	84	16	0.99	0.15	99	1	83.97	72.00
76	C00097	84	76	0.98	0.35	98	2	84.23	65.00
77	C00098	84	77	0.81	0.17	81	19	84.48	81.16

11	C00031	83	11	0.98	0.18	98	2	84.05	74.00
26	C00046	83	26	0.65	0.14	65	35	84.63	82.40
33	C00053	83	33	0.65	0.14	65	35	84.63	82.40
41	C00061	83	41	0.97	0.10	97	3	83.99	79.33
44	C00064	83	44	0.65	0.14	65	35	84.63	82.40
54	C00074	83	54	0.38	0.33	38	62	87.16	81.82
55	C00075	83	55	0.80	0.47	80	20	85.68	76.55
57	C00077	83	57	0.78	0.46	78	22	85.73	77.18
63	C00083	83	63	0.98	0.13	98	2	83.99	77.00
66	C00086	83	66	0.98	0.35	98	2	84.23	65.00
68	C00088	83	68	0.98	0.35	98	2	84.23	65.00
73	C00093	83	73	0.76	0.31	76	24	85.18	79.63
96	C00117	83	96	0.99	0.15	99	1	83.97	72.00
99	C00120	83	99	0.98	0.35	98	2	84.23	65.00
24	C00044	82	24	0.99	0.26	99	1	84.05	64.00
34	C00054	82	34	0.99	0.26	99	1	84.05	64.00
35	C00055	82	35	0.98	0.20	98	2	84.07	73.00
38	C00058	82	38	0.99	0.15	99	1	83.97	72.00
39	C00059	82	39	0.65	0.14	65	35	84.63	82.40
48	C00068	82	48	0.97	0.31	97	3	84.28	70.00
50	C00070	82	50	0.76	0.32	76	24	85.24	79.46
51	C00071	82	51	0.65	0.14	65	35	84.63	82.40
52	C00072	82	52	0.98	0.35	98	2	84.23	65.00
61	C00081	82	61	0.98	0.13	98	2	83.99	77.00
67	C00087	82	67	0.99	0.15	99	1	83.97	72.00
69	C00089	82	69	0.81	0.17	81	19	84.48	81.16
71	C00091	82	71	0.97	0.22	97	3	84.14	74.33
72	C00092	82	72	0.40	0.34	40	60	87.03	81.73
74	C00094	82	74	0.98	0.35	98	2	84.23	65.00
79	C00100	82	79	0.98	0.35	98	2	84.23	65.00
84	C00105	82	84	0.98	0.35	98	2	84.23	65.00
90	C00111	82	90	0.75	0.31	75	25	85.24	79.68
93	C00114	82	93	0.99	0.15	99	1	83.97	72.00
94	C00115	82	94	0.77	0.35	77	23	85.35	78.83
4	C00024	81	4	0.81	0.53	81	19	85.83	75.42
5	C00025	81	5	0.92	0.12	92	8	84.13	80.63
12	C00032	81	12	0.67	0.14	67	33	84.61	82.30
29	C00049	81	29	0.98	0.35	98	2	84.23	65.00
58	C00078	81	58	0.38	0.33	38	62	87.16	81.82
9	C00029	80	9	0.82	0.13	82	18	84.33	81.67
18	C00038	80	18	0.65	0.14	65	35	84.63	82.40
86	C00107	80	86	0.98	0.35	98	2	84.23	65.00
8	C00028	79	8	0.78	0.37	78	22	85.38	78.41
17	C00037	78	17	0.81	0.14	81	19	84.37	81.63
22	C00042	78	22	0.78	0.41	78	22	85.54	77.86

20	C00040	77	20	0.39	0.30	39	61	86.74	82.00
95	C00116	76	95	0.98	0.35	98	2	84.23	65.00
7	C00027	75	7	0.99	0.15	99	1	83.97	72.00
15	C00035	75	15	0.98	0.12	98	2	83.98	77.50
36	C00056	74	36	0.77	0.36	77	23	85.36	78.78
60	C00080	74	60	0.96	0.17	96	4	84.13	77.25
13	C00033	72	13	0.98	0.11	98	2	83.97	78.00
3	C00023	71	3	0.67	0.14	67	33	84.61	82.30
6	C00026	71	6	0.79	0.34	79	21	85.20	78.76
10	C00030	70	10	0.98	0.20	98	2	84.07	73.00
19	C00039	70	19	0.99	0.15	99	1	83.97	72.00
27	C00047	70	27	0.97	0.25	97	3	84.20	72.67
30	C00050	70	30	0.98	0.35	98	2	84.23	65.00
65	C00085	70	65	0.81	0.17	81	19	84.48	81.16
82	C00103	70	82	0.99	0.15	99	1	83.97	72.00
81	C00102	66	81	0.98	0.35	98	2	84.23	65.00
56	C00076	64	56	0.65	0.14	65	35	84.63	82.40

## Appendix E: Item difficulty (*P*) Values Ranked and Ordered

Examinees Scores			Item Statistics						
Examinees	ID	Score	Item	<i>P</i>	<i>Rpb</i>	Number Correct	Number Incorrect	Mean score correct	Mean score incorrect
88	C00109	94	88	0.99	0.15	99	1	83.97	72.00
45	C00065	93	45	0.99	0.15	99	1	83.97	72.00
16	C00036	84	16	0.99	0.15	99	1	83.97	72.00
96	C00117	83	96	0.99	0.15	99	1	83.97	72.00
24	C00044	82	24	0.99	0.26	99	1	84.05	64.00
34	C00054	82	34	0.99	0.26	99	1	84.05	64.00
38	C00058	82	38	0.99	0.15	99	1	83.97	72.00
67	C00087	82	67	0.99	0.15	99	1	83.97	72.00
93	C00114	82	93	0.99	0.15	99	1	83.97	72.00
7	C00027	75	7	0.99	0.15	99	1	83.97	72.00
19	C00039	70	19	0.99	0.15	99	1	83.97	72.00
82	C00103	70	82	0.99	0.15	99	1	83.97	72.00
42	C00062	94	42	0.98	0.35	98	2	84.23	65.00
92	C00113	94	92	0.98	0.35	98	2	84.23	65.00
43	C00063	93	43	0.98	0.35	98	2	84.23	65.00
89	C00110	93	89	0.98	0.35	98	2	84.23	65.00
100	C00121	92	100	0.98	0.35	98	2	84.23	65.00
21	C00041	85	21	0.98	0.35	98	2	84.23	65.00
76	C00097	84	76	0.98	0.35	98	2	84.23	65.00
11	C00031	83	11	0.98	0.18	98	2	84.05	74.00
63	C00083	83	63	0.98	0.13	98	2	83.99	77.00
66	C00086	83	66	0.98	0.35	98	2	84.23	65.00
68	C00088	83	68	0.98	0.35	98	2	84.23	65.00
99	C00120	83	99	0.98	0.35	98	2	84.23	65.00
35	C00055	82	35	0.98	0.20	98	2	84.07	73.00
52	C00072	82	52	0.98	0.35	98	2	84.23	65.00
61	C00081	82	61	0.98	0.13	98	2	83.99	77.00
74	C00094	82	74	0.98	0.35	98	2	84.23	65.00
79	C00100	82	79	0.98	0.35	98	2	84.23	65.00
84	C00105	82	84	0.98	0.35	98	2	84.23	65.00
29	C00049	81	29	0.98	0.35	98	2	84.23	65.00
86	C00107	80	86	0.98	0.35	98	2	84.23	65.00
95	C00116	76	95	0.98	0.35	98	2	84.23	65.00
15	C00035	75	15	0.98	0.12	98	2	83.98	77.50
13	C00033	72	13	0.98	0.11	98	2	83.97	78.00
10	C00030	70	10	0.98	0.20	98	2	84.07	73.00
30	C00050	70	30	0.98	0.35	98	2	84.23	65.00
81	C00102	66	81	0.98	0.35	98	2	84.23	65.00

98	C00119	95	98	0.97	0.19	97	3	84.10	75.67
91	C00112	92	91	0.97	0.13	97	3	84.03	78.00
46	C00066	90	46	0.97	0.22	97	3	84.15	74.00
59	C00079	90	59	0.97	0.35	97	3	84.33	68.33
41	C00061	83	41	0.97	0.10	97	3	83.99	79.33
48	C00068	82	48	0.97	0.31	97	3	84.28	70.00
71	C00091	82	71	0.97	0.22	97	3	84.14	74.33
27	C00047	70	27	0.97	0.25	97	3	84.20	72.67
2	C00022	84	2	0.96	0.19	96	4	84.16	76.50
60	C00080	74	60	0.96	0.17	96	4	84.13	77.25
5	C00025	81	5	0.92	0.12	92	8	84.13	80.63
1	C00021	88	1	0.83	0.13	83	17	84.31	81.59
9	C00029	80	9	0.82	0.13	82	18	84.33	81.67
25	C00045	96	25	0.81	0.17	81	19	84.48	81.16
80	C00101	96	80	0.81	0.17	81	19	84.48	81.16
85	C00106	94	85	0.81	0.17	81	19	84.48	81.16
75	C00096	92	75	0.81	0.17	81	19	84.48	81.16
14	C00034	89	14	0.81	0.14	81	19	84.37	81.63
77	C00098	84	77	0.81	0.17	81	19	84.48	81.16
69	C00089	82	69	0.81	0.17	81	19	84.48	81.16
4	C00024	81	4	0.81	0.53	81	19	85.83	75.42
17	C00037	78	17	0.81	0.14	81	19	84.37	81.63
65	C00085	70	65	0.81	0.17	81	19	84.48	81.16
87	C00108	95	87	0.80	0.47	80	20	85.68	76.55
32	C00052	92	32	0.80	0.47	80	20	85.65	76.65
55	C00075	83	55	0.80	0.47	80	20	85.68	76.55
6	C00026	71	6	0.79	0.34	79	21	85.20	78.76
53	C00073	95	53	0.78	0.47	78	22	85.77	77.05
47	C00067	94	47	0.78	0.39	78	22	85.46	78.14
57	C00077	83	57	0.78	0.46	78	22	85.73	77.18
8	C00028	79	8	0.78	0.37	78	22	85.38	78.41
22	C00042	78	22	0.78	0.41	78	22	85.54	77.86
64	C00084	96	64	0.77	0.33	77	23	85.25	79.17
28	C00048	92	28	0.77	0.35	77	23	85.35	78.83
94	C00115	82	94	0.77	0.35	77	23	85.35	78.83
36	C00056	74	36	0.77	0.36	77	23	85.36	78.78
62	C00082	95	62	0.76	0.36	76	24	85.43	78.83
70	C00090	93	70	0.76	0.28	76	24	85.05	80.04
73	C00093	83	73	0.76	0.31	76	24	85.18	79.63
50	C00070	82	50	0.76	0.32	76	24	85.24	79.46
40	C00060	92	40	0.75	0.28	75	25	85.09	80.12
90	C00111	82	90	0.75	0.31	75	25	85.24	79.68
78	C00099	94	78	0.74	0.28	74	26	85.14	80.19
97	C00118	94	97	0.74	0.30	74	26	85.24	79.88
83	C00104	93	83	0.74	0.30	74	26	85.24	79.88

12	C00032	81	12	0.67	0.14	67	33	84.61	82.30
3	C00023	71	3	0.67	0.14	67	33	84.61	82.30
23	C00043	88	23	0.65	0.14	65	35	84.63	82.40
26	C00046	83	26	0.65	0.14	65	35	84.63	82.40
33	C00053	83	33	0.65	0.14	65	35	84.63	82.40
44	C00064	83	44	0.65	0.14	65	35	84.63	82.40
39	C00059	82	39	0.65	0.14	65	35	84.63	82.40
51	C00071	82	51	0.65	0.14	65	35	84.63	82.40
18	C00038	80	18	0.65	0.14	65	35	84.63	82.40
56	C00076	64	56	0.65	0.14	65	35	84.63	82.40
72	C00092	82	72	0.40	0.34	40	60	87.03	81.73
31	C00051	93	31	0.39	0.34	39	61	87.10	81.77
20	C00040	77	20	0.39	0.30	39	61	86.74	82.00
49	C00069	95	49	0.38	0.33	38	62	87.16	81.82
54	C00074	83	54	0.38	0.33	38	62	87.16	81.82
58	C00078	81	58	0.38	0.33	38	62	87.16	81.82
37	C00057	91	37	0.35	0.26	35	65	86.63	82.35



## Appendix F: Discrimination Values ( $r_{pb}$ ) Ranked and Ordered

### Examinees Scores

### Item Statistics

Examinees	ID	Score	Item	$P$	$R_{pb}$	Number Correct	Number Incorrect	Mean score correct	Mean score incorrect
4	C00024	81	4	0.81	0.53	81	19	85.83	75.42
87	C00108	95	87	0.80	0.47	80	20	85.68	76.55
32	C00052	92	32	0.80	0.47	80	20	85.65	76.65
55	C00075	83	55	0.80	0.47	80	20	85.68	76.55
53	C00073	95	53	0.78	0.47	78	22	85.77	77.05
57	C00077	83	57	0.78	0.46	78	22	85.73	77.18
22	C00042	78	22	0.78	0.41	78	22	85.54	77.86
47	C00067	94	47	0.78	0.39	78	22	85.46	78.14
8	C00028	79	8	0.78	0.37	78	22	85.38	78.41
36	C00056	74	36	0.77	0.36	77	23	85.36	78.78
62	C00082	95	62	0.76	0.36	76	24	85.43	78.83
42	C00062	94	42	0.98	0.35	98	2	84.23	65.00
92	C00113	94	92	0.98	0.35	98	2	84.23	65.00
43	C00063	93	43	0.98	0.35	98	2	84.23	65.00
89	C00110	93	89	0.98	0.35	98	2	84.23	65.00
100	C00121	92	100	0.98	0.35	98	2	84.23	65.00
21	C00041	85	21	0.98	0.35	98	2	84.23	65.00
76	C00097	84	76	0.98	0.35	98	2	84.23	65.00
66	C00086	83	66	0.98	0.35	98	2	84.23	65.00
68	C00088	83	68	0.98	0.35	98	2	84.23	65.00
99	C00120	83	99	0.98	0.35	98	2	84.23	65.00
52	C00072	82	52	0.98	0.35	98	2	84.23	65.00
74	C00094	82	74	0.98	0.35	98	2	84.23	65.00
79	C00100	82	79	0.98	0.35	98	2	84.23	65.00
84	C00105	82	84	0.98	0.35	98	2	84.23	65.00
29	C00049	81	29	0.98	0.35	98	2	84.23	65.00
86	C00107	80	86	0.98	0.35	98	2	84.23	65.00
95	C00116	76	95	0.98	0.35	98	2	84.23	65.00
30	C00050	70	30	0.98	0.35	98	2	84.23	65.00
81	C00102	66	81	0.98	0.35	98	2	84.23	65.00
59	C00079	90	59	0.97	0.35	97	3	84.33	68.33
28	C00048	92	28	0.77	0.35	77	23	85.35	78.83
94	C00115	82	94	0.77	0.35	77	23	85.35	78.83
6	C00026	71	6	0.79	0.34	79	21	85.20	78.76
72	C00092	82	72	0.40	0.34	40	60	87.03	81.73
31	C00051	93	31	0.39	0.34	39	61	87.10	81.77
64	C00084	96	64	0.77	0.33	77	23	85.25	79.17
49	C00069	95	49	0.38	0.33	38	62	87.16	81.82
54	C00074	83	54	0.38	0.33	38	62	87.16	81.82

58	C00078	81	58	0.38	0.33	38	62	87.16	81.82
50	C00070	82	50	0.76	0.32	76	24	85.24	79.46
48	C00068	82	48	0.97	0.31	97	3	84.28	70.00
73	C00093	83	73	0.76	0.31	76	24	85.18	79.63
90	C00111	82	90	0.75	0.31	75	25	85.24	79.68
97	C00118	94	97	0.74	0.30	74	26	85.24	79.88
83	C00104	93	83	0.74	0.30	74	26	85.24	79.88
20	C00040	77	20	0.39	0.30	39	61	86.74	82.00
70	C00090	93	70	0.76	0.28	76	24	85.05	80.04
40	C00060	92	40	0.75	0.28	75	25	85.09	80.12
78	C00099	94	78	0.74	0.28	74	26	85.14	80.19
24	C00044	82	24	0.99	0.26	99	1	84.05	64.00
34	C00054	82	34	0.99	0.26	99	1	84.05	64.00
37	C00057	91	37	0.35	0.26	35	65	86.63	82.35
27	C00047	70	27	0.97	0.25	97	3	84.20	72.67
46	C00066	90	46	0.97	0.22	97	3	84.15	74.00
71	C00091	82	71	0.97	0.22	97	3	84.14	74.33
35	C00055	82	35	0.98	0.20	98	2	84.07	73.00
10	C00030	70	10	0.98	0.20	98	2	84.07	73.00
98	C00119	95	98	0.97	0.19	97	3	84.10	75.67
2	C00022	84	2	0.96	0.19	96	4	84.16	76.50
11	C00031	83	11	0.98	0.18	98	2	84.05	74.00
60	C00080	74	60	0.96	0.17	96	4	84.13	77.25
25	C00045	96	25	0.81	0.17	81	19	84.48	81.16
80	C00101	96	80	0.81	0.17	81	19	84.48	81.16
85	C00106	94	85	0.81	0.17	81	19	84.48	81.16
75	C00096	92	75	0.81	0.17	81	19	84.48	81.16
77	C00098	84	77	0.81	0.17	81	19	84.48	81.16
69	C00089	82	69	0.81	0.17	81	19	84.48	81.16
65	C00085	70	65	0.81	0.17	81	19	84.48	81.16
88	C00109	94	88	0.99	0.15	99	1	83.97	72.00
45	C00065	93	45	0.99	0.15	99	1	83.97	72.00
16	C00036	84	16	0.99	0.15	99	1	83.97	72.00
96	C00117	83	96	0.99	0.15	99	1	83.97	72.00
38	C00058	82	38	0.99	0.15	99	1	83.97	72.00
67	C00087	82	67	0.99	0.15	99	1	83.97	72.00
93	C00114	82	93	0.99	0.15	99	1	83.97	72.00
7	C00027	75	7	0.99	0.15	99	1	83.97	72.00
19	C00039	70	19	0.99	0.15	99	1	83.97	72.00
82	C00103	70	82	0.99	0.15	99	1	83.97	72.00
14	C00034	89	14	0.81	0.14	81	19	84.37	81.63
17	C00037	78	17	0.81	0.14	81	19	84.37	81.63
12	C00032	81	12	0.67	0.14	67	33	84.61	82.30
3	C00023	71	3	0.67	0.14	67	33	84.61	82.30
23	C00043	88	23	0.65	0.14	65	35	84.63	82.40

26	C00046	83	26	0.65	0.14	65	35	84.63	82.40
33	C00053	83	33	0.65	0.14	65	35	84.63	82.40
44	C00064	83	44	0.65	0.14	65	35	84.63	82.40
39	C00059	82	39	0.65	0.14	65	35	84.63	82.40
51	C00071	82	51	0.65	0.14	65	35	84.63	82.40
18	C00038	80	18	0.65	0.14	65	35	84.63	82.40
56	C00076	64	56	0.65	0.14	65	35	84.63	82.40
63	C00083	83	63	0.98	0.13	98	2	83.99	77.00
61	C00081	82	61	0.98	0.13	98	2	83.99	77.00
91	C00112	92	91	0.97	0.13	97	3	84.03	78.00
1	C00021	88	1	0.83	0.13	83	17	84.31	81.59
9	C00029	80	9	0.82	0.13	82	18	84.33	81.67
15	C00035	75	15	0.98	0.12	98	2	83.98	77.50
5	C00025	81	5	0.92	0.12	92	8	84.13	80.63
13	C00033	72	13	0.98	0.11	98	2	83.97	78.00
41	C00061	83	41	0.97	0.10	97	3	83.99	79.33

## Appendix G: Distractor Frequencies and Mean Scores

Item	Key	Frequencies				Mean Score			
		A	B	C	D	A	B	C	D
1	A	82	18	0	0	84.33	81.67		
2	A	96	1	1	2	84.16	80.00	74.00	76.00
3	B	0	67	31	2		84.61	83.42	65.00
4	C	0	16	81	3		75.94	85.83	72.67
5	A	92	4	2	2	84.13	82.50	77.50	80.00
6	C	1	18	79	2	92.00	79.00	85.20	70.00
7	A	99	0	0	1	83.97			72.00
8	C	0	19	78	3		79.32	85.38	72.67
9	A	82	18	0	0	84.33	81.67		
10	A	98	0	1	1	84.07		74.00	72.00
11	A	98	1	0	1	84.05	76.00		72.00
12	B	0	67	31	2		84.61	83.42	65.00
13	A	98	0	1	1	83.97		81.00	75.00
14	A	81	19	0	0	84.37	81.63		
15	A	98	1	0	1	83.98	74.00		81.00
16	A	99	0	0	1	83.97			72.00
17	A	81	19	0	0	84.37	81.63		
18	B	0	65	33	2		84.63	83.45	65.00
19	A	99	0	0	1	83.97			72.00
20	D	0	0	61	39			82.00	86.74
21	D	0	0	2	98			65.00	84.23
22	C	0	19	78	3		78.68	85.54	72.67
23	B	0	65	33	2		84.63	83.45	65.00
24	D	0	0	1	99			64.00	84.05
25	A	81	18	0	1	84.48	81.67		72.00
26	B	0	65	33	2		84.63	83.45	65.00
27	D	0	0	3	97			72.67	84.20
28	C	0	20	77	3		79.75	85.35	72.67
29	D	0	0	2	98			65.00	84.23
30	D	0	0	2	98			65.00	84.23
31	D	0	0	61	39			81.77	87.10
32	C	0	15	80	5		74.93	85.65	81.80
33	B	0	65	33	2		84.63	83.45	65.00
34	D	0	0	1	99			64.00	84.05
35	A	98	0	1	1	84.07		74.00	72.00
36	C	0	20	77	3		79.70	85.36	72.67
37	D	0	0	65	35			82.35	86.63
38	A	99	0	0	1	83.97			72.00
39	B	0	65	33	2		84.63	83.45	65.00
40	C	0	22	75	3		81.14	85.09	72.67
41	A	97	0	0	3	83.99			79.33
42	D	0	0	2	98			65.00	84.23
43	D	0	0	2	98			65.00	84.23
44	B	0	65	33	2		84.63	83.45	65.00
45	A	99	0	0	1	83.97			72.00
46	D	0	0	3	97			74.00	84.15
47	C	0	19	78	3		79.00	85.46	72.67

48	D	0	0	3	97		70.00	84.28
49	D	0	0	62	38		81.82	87.16
50	C	0	21	76	3	80.43	85.24	72.67
51	B	0	65	33	2	84.63	83.45	65.00
52	D	0	0	2	98		65.00	84.23
53	C	0	15	78	7	74.00	85.77	74.14
54	D	0	0	62	38		81.82	87.16
55	C	0	17	80	3	77.24	48.27	72.67
56	B	0	65	33	2	84.63	83.45	65.00
57	C	0	19	78	3	77.89	85.73	72.67
58	D	0	0	62	38		81.82	87.16
59	D	0	0	3	97		68.33	84.33
60	A	96	0	3	1	84.13	79.00	81.00
61	A	98	0	1	1	83.99	82.00	72.00
62	C	0	18	76	6	77.89	85.43	81.67
63	A	98	0	0	2	83.99		77.00
64	C	0	16	77	7	77.00	85.25	84.14
65	A	81	18	0	1	84.48	81.67	72.00
66	D	0	0	2	98		65.00	84.23
67	A	99	0	0	1	83.97		72.00
68	D	0	0	2	98		65.00	84.23
69	A	81	18	0	1	84.48	81.67	72.00
70	C	0	13	76	11	73.15	85.05	88.18
71	D	0	0	3	97		74.33	84.14
72	D	0	0	60	40		81.73	87.03
73	C	0	18	76	6	78.22	85.18	83.83
74	D	0	0	2	98		65.00	84.23
75	A	81	18	0	1	84.48	81.67	72.00
76	D	0	0	2	98		65.00	84.23
77	A	81	18	0	1	84.48	81.67	72.00
78	C	0	23	74	3	81.17	85.14	72.67
79	D	0	0	2	98		65.00	84.23
80	A	81	18	0	1	84.48	81.67	72.00
81	D	0	0	2	98		65.00	84.23
82	A	99	0	0	1	83.97		72.00
83	D	0	23	74	3	80.83	85.24	72.67
84	D	0	0	2	98		65.00	84.23
85	A	81	18	0	1	84.48	81.67	72.00
86	D	0	0	2	98		65.00	84.23
87	C	0	17	80	3	77.24	85.68	72.67
88	A	99	0	0	1	83.97		72.00
89	D	0	0	2	98		65.00	84.23
90	C	0	22	75	3	80.64	85.24	72.67
91	A	97	0	2	1	84.03	81.00	72.00
92	D	0	0	2	98		65.00	84.23
93	A	99	0	0	1	83.97		72.00
94	C	0	18	77	5	78.17	85.35	81.20
95	D	0	0	2	98		65.00	84.23
96	A	99	0	0	1	83.97		72.00
97	C	0	18	74	8	78.17	85.24	83.75
98	A	97	2	0	1	84.10	77.50	72.00
99	D	0	0	2	98		65.00	84.23
100	D	0	0	2	98		65.00	84.23

## References

1. American Educational Research Association (AERA), American Psychological Association (APA), National Council on Measurement in Education (NCME): *Standards for Educational and Psychological Testing* (1999).
2. Clauser, B.E., Margolis, M.J., & Case, S.M. (2006). Testing for licensure and certification in the professions. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 701-731). Westport, CT: Praeger.
3. Gene V. Glass and Kenneth D. Hopkins (1995), *Statistical Methods in Education and Psychology* (3rd edition).
4. Kane, M. (2004). Certification testing as an illustration of argument- based validation. *Measurement*, 2, 135-170.
5. Kaplan and Saccuzzo. *Psychological Testing: Principles, Application, and Issues*. p 113, KR-20-21 formulas. KR-20-21.
6. Knapp, Joan, Anderson, Lynn, Wild, Cheryl: Institute for Credentialing Excellence (ICE): *The ICE Handbook*, (2004).
7. Kuder, G. F.; M. W. Richardson (1937) the theory of the estimation of test reliability. *Psychometrika*, 2: 151-60.
8. Mislevy, R.J., Steinberg, L.S., Breyer, F.J., Almond, P.O., & Johnson, L. (2002). Making sense of data from complex assessments. *Applied Measurement in Education*, 15, 363-389.
9. Mehrens, WA. (1995). Legal and professional bases for licensure testing. In iC. Impara (Ed.), *Licensure Testing: Purposes, procedures, and practices* (pp. 35-58). Lincoln, NE: **Buros Institute of Mental Measurements**.
10. National Commission for Certifying Agencies (NCCA): *Standards for the Accreditation of Certification Programs*, (2004). Chapter 10: The certification program must analyze, define, and publish performance domains and tasks related to the purpose of the credential, and the knowledge and or skills associated with the performance domains and tasks, and use them to develop specifications for the assessment instrument.
11. National Commission for Certifying Agencies (NCCA): *Standards for the Accreditation of Certification Programs*, (2004). Chapter 11: The certification program must employ assessment instruments that are derived from the job/practice analysis and that are consistent with generally accepted psychometric principles.
12. National Commission for Certifying Agencies (NCCA): *Standards for the Accreditation of Certification Programs*, (2004). Chapter 12: The Certification program must set the cut score

consistent with the purpose of the credential and the established standards of competence for profession, occupation, role, or skill.

13. Shultz, K.S., & Whitney, D.J. (2005) *Measurement theory in action*. Thousand Oaks, CA: Sage. BF 39 .S55 2005.
14. Steven G. Rogerlberg. *Handbook of research methods in industrial organizational psychology*. p 327, Rpbis Formulas. Rpbis.
15. Raymond, MR. (1996). Establishing weights for test plans for licensure and certification examinations. *Applied Measurement in Education*, 9, 237-256.
16. Raymond, MR. (2001). Job analysis and the specification of content for licensure and certification examinations. *Applied Measurement in Education*, 14, 369-415.
17. Raymond, MR. (2002). A practical guide to practice analysis for credentialing examinations. *Educational Measurement: Issues and Practice*, 21, 25-37.
18. Raymond, MR., & Neustel, S. Determining the content of credentialing examinations. In S.M. Downing & TM. Haladyna (Eds.), *Handbook of Test Development* (pp. 181-224). Mahwah, NJ: Lawrence Erlbaum.
19. Schmitt, K. (1995). What is licensure? In J.C. Impara (Ed.), *Licensure testing: Purposes, procedures, and practices* (pp. 3-32). Lincoln, NE: Buros Institute of Mental Measurements.
20. Tannenbaum, R.J. (1999). Laying the groundwork for a licensure assessment. *Journal of Personnel Evaluation in Education*, 13, 225-244.
21. *Uniform Guidelines on Employee Selection Procedures*, Equal Employment Opportunity Commission, Civil Service Commission, US Department of Labor, US Department of Justice, 1978.
22. Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement*, 21, 361-375.