



Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2010

Utility of the General Validity Scale Model: Development of Validity Scales for the Co- parenting Behavior Questionnaire

Kimberly Parker

Virginia Commonwealth University

Follow this and additional works at: <http://scholarscompass.vcu.edu/etd>

 Part of the [Psychology Commons](#)

© The Author

Downloaded from

<http://scholarscompass.vcu.edu/etd/2301>

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

UTILITY OF THE GENERAL VALIDITY SCALE MODEL: DEVELOPMENT OF
VALIDITY SCALES FOR THE CO-PARENTING BEHAVIOR QUESTIONNAIRE

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science
at Virginia Commonwealth University.

By: KIMBERLY M. PARKER
Bachelor of Arts
University of North Carolina at Chapel Hill, 2006

Director: Dr. Arnold L. Stolberg
Emeritus Professor of Psychology

Virginia Commonwealth University
Richmond, Virginia
December, 2010

Acknowledgements

I extend a gracious thank you to each of my committee members. Thank you to Dr. Robert Archer, for his knowledge and expertise in the area of validity scales. Thank you to Dr. Bryce McLeod, for his suggestions, guidance, and extra push to help further develop my research. I especially thank my thesis director, Dr. Arnold Stolberg, for his direction, knowledge, support, outstanding dedication, and openness to the goals and direction of his students. Finally, I thank my colleagues, friends, and family for their understanding and support in keeping me grounded and balanced through my journey.

Table of Contents

Acknowledgements	ii
List of Tables	iv
Abstract	v
Introduction	1
Common Threats to Validity	4
Current Validity Scales Detecting Validity Threats	5
General Validity Scale Model	14
Application of the General Validity Scale Model to the CBQ	21
Hypotheses	27
Study 1	29
Method	29
Results	34
Discussion	46
Study 2	50
Method	50
Results	58
Discussion	63
General Discussion	67
References	77
Appendices	79
A: Item Loadings of the CBQ by Subscale	79
B: Co-parenting Behavior Questionnaire - Revised	80
C: Student Demographic Questionnaire	92
Vita	94

List of Tables

	Page
Table 1. Existing Measures and Procedures for Validity Scale Development	10
Table 2. Examples of Unresolved Threats to Validity.....	13
Table 3. Steps of the General Validity Scale Model.....	14
Table 4. Threats to Validity, Threat Subtypes, and Procedural Solutions for Addressing Threats	18
Table 5. Application of the General Validity Scale Model to the CBQ	22
Table 6. CBQ Items, Frequencies, and Subscale Loadings for NPP Scales.....	36
Table 7. Finalized Item Inclusion for NPP Scales	38
Table 8. CBQ Items, Frequencies, and Subscale Loadings for PPP Scales.....	40
Table 9. Item Pairs, Frequencies, and Subscale Loading for the Parenting Discrepancy Scale.	42
Table 10. Item Pairs for Inconsistency Scale.....	44
Table 11. Reading Comprehension Items and Grade Level	45
Table 12. Results of T-tests to Validate Each Validity Scale	60

Abstract

UTILITY OF THE GENERAL VALIDITY SCALE MODEL: DEVELOPMENT OF VALIDITY SCALES FOR THE CO-PARENTING BEHAVIOR QUESTIONNAIRE

Kimberly M. Parker, B.A.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science
at Virginia Commonwealth University.

Virginia Commonwealth University, 2010

Director: Dr. Arnold L. Stolberg
Emeritus Professor of Psychology

Validity scales for child-report measures are necessary tools in clinical and forensic settings in which major decisions affecting the child and family are in question. Currently there is no standard model for the development and testing of such validity scales. The present study focused on 1) creating the General Validity Scale (GVS) Model to serve as a guide in validity scale development and 2) applying this model in the development of validity scales for the Co-parenting Behavior Questionnaire (CBQ), a child-report measure of parenting and co-parenting behaviors for children whose parents are divorced. Study 1 used the newly developed GVS Model to identify threats to CBQ validity and to develop procedures for detecting such threats. Four different validity scales were created to detect inaccurate responding due to 1) presenting mothering, fathering, and/or co-parenting in an overly negative light, 2) rating mothering and fathering in a highly discrepant manner, 3) inconsistent item responses, and 4) low reading level.

Study 2 followed the GVS Model to test the newly developed scales by comparing CBQ

responses produced under a standard instruction set to responses from contrived or randomly generated data. Support for the ability of each validity scale to accurately detect threats to validity was found.

Utility of the General Validity Scale Model: Development of Validity Scales for the Co-parenting Behavior Questionnaire

Accurately measuring parenting is a difficult and important clinical task, especially when divorce is part of the family picture. The primary task is to gain accurate measurements of the target behaviors. Divorce and its myriad of complex processes complicate the assessment task. Of greatest interest to this study is the influence and assessment of efforts to intentionally or inadvertently misrepresent the parenting behaviors in question. A General Validity Scale Model is proposed as a guideline for addressing invalid measurement, and includes identifying threats to validity, proposing possible scales that may assess validity, and developing techniques used in the construction of such scales. This General Validity Scale Model will be applied in the first steps of development for validity scales of the Co-parenting Behavior Questionnaire, an assessment tool of use in divorce and custody evaluations.

Intentional efforts to manipulate test outcomes and random responding are among reasons individuals have distorted responses in assessment (Piedmont, 1998). When these and other threats to validity surface, validity scales in formal assessment procedures may serve to alert test evaluators of response sets and biases that invalidate measurement. Validity scales are selected items embedded in a complete measure that may detect the extent to which a test taker is responding in a biased manner. Once biased responses are detected, the test administrator may opt to do one of several things including: 1) identify the test as invalid and disregard the test scores, 2) retain the scores but use them with caution, or 3) use validity scales to identify psychopathological influences that may operate in interfering with accurate responding (Furr & Bacharach, 2007). Although validity scales promise great utility, very

few psychological measures, particularly child report measures, include scales to evaluate validity of responses.

Validity scales prove particularly important in contexts in which clinical, forensic, or employment decisions must be made (Edens & Ruiz, 2006; Piedmont, 1998). Failure to detect an invalid protocol in a clinical setting may lead to misdiagnosing, failing to treat those in need, or providing the wrong treatment to clients. Using invalid data in forensic evaluations may result in inaccurate determinations of guilt, misclassifications of sexual abuse, or unhealthy assignments of child custody. In job settings, validity scales may be useful in identifying applicants who attempt to give unrealistically positive impressions of themselves in hopes of filling a position or gaining a promotion (Butcher, 1979; Baer & Wetter, 1997).

Using validity scales may be especially useful in the clinical and forensic settings when assessing families of divorce (Baer & Wetter, 1997). Decisions made during custody evaluations can have a profound impact on the family dynamic and on a child's wellbeing. However, parents and children alike may have several reasons for providing inaccurate and unreliable information to psychologists and courts. For example, parents will present themselves positively and the other parent negatively in hopes of being awarded custody of their child(ren). Children may want to stay out of the divorce as much of possible and will, therefore, provide little or neutral information, or they may present one parent favorably for reasons that may not be in their best interest. Although ethical standards require evaluators to use multiple, valid sources of information when making custody recommendations, measures with validity scales imbedded will be useful to the custody evaluation process.

The Co-parenting Behavior Questionnaire (CBQ) is a tool used to assess parenting and co-parenting behaviors in child custody evaluations and in psychotherapy cases in which separation and divorce figure prominently (Mullet & Stolberg, 1999). The CBQ is an 86-item questionnaire that assesses twelve dimensions of post-divorce parenting and co-parenting from the child's perspective. The twelve subscales of the CBQ measure the 4 co-parenting variables of Conflict, Triangulation, Parental Respect/Cooperation, and Co-parental Communication as well as the 4 mother-parenting and 4 father-parenting behaviors of Parental Warmth/Acceptance, Discipline, Monitoring, and Parent-Child Communication. These domains have been considered parenting and co-parenting factors salient to child adjustment and maladjustment. Cronbach's alpha for the scales of the CBQ range from .82 to .93 suggesting adequate internal consistency of the measure. The Co-parenting Behavior Questionnaire has been normed on a sample of children aged 10 to 18. Evidence of the predictive validity of the CBQ has been established as its scores are correlated with several important dimensions of child adjustment including total behavior problems, self-esteem, acting out, antisocial behavior, headstrong behavior, and anxiety/depression (Macie & Stolberg, 2003). The CBQ holds promise in identifying maladaptive parenting and co-parenting strategies that may guide treatments for post-divorce parents and families with the possibility of use as an assessment tool to monitor treatment outcomes over time. As a measure of parenting and co-parenting the CBQ shows utility in divorce and custody evaluations in which such behaviors are in question. Because of its use as a decision-making tool in forensic settings, the need exists to create empirically derived scales that will identify random and biased response-patterns.

Common Threats to Validity

Establishing a strategy to develop validity scales for a specific measure involves three key steps. First, it is important to identify the threats to accurate responding that are relevant for the measure and the domains that are being assessed. Next, existing or novel strategies to measure these specific threats to validity must be identified. Finally, these possible validity scales must be tested and normed.

Common sources that may compromise the reliability or validity within any given scale of measurement include random or inconsistent responding and over- and under-reporting of problems. **Random responding** may present as an issue if, for example, the respondent has a low reading comprehension level, if they are confused because of emotional or cognitive dysfunction, or if they are disengaged or uninterested in completing the task. If a measure has a long administration time, like the MMPI-2, inconsistent responding may occur as the test taker becomes fatigued over time. **Over-reporting of symptoms or problem behaviors** may arise if the respondent feels as if they are worse off than they actually are, as in the case of elevated F scales on the MMPI-2 (Greene, 2000). Others may intentionally try to make themselves or another individual appear worse off in order to gain benefits or avoid punishment. For example, one might over-report problems if they are seeking disability services that must meet certain criteria or a criminal may exaggerate psychopathology to qualify for an insanity defense and, thus, avoid a severe punishment. Invalid responding may also occur for the very opposite reason, when the respondent **underreports problems**. An individual may underreport if they are defensive or in denial about the existence of their problems (Greene, 2000). If a person experiences substantial depression, for example, they may underreport symptoms to avoid the stigma attached to this mental illness.

Another threat to validity that may occur only in select instruments is unusually large discrepancies in appraisals of the behaviors or symptoms being assessed at one time. These *large discrepancies in responses* may pose as a threat to validity when respondents are reporting on multiple people or behaviors within a measure. For example, if a parent is reporting on behaviors for several biological children, abnormally high scores for one child and unusually low behavior scores for another child may indicate an over- or under-exaggeration of responses and thus indicate invalidity. This form of invalidity has not been examined to date in any instrument nor are there existing scales to detect this type of responding. Test developers are however often aware of the former three threats to validity, and in fact, current validity scales commonly assess for these possible reasons of invalid data (Greene, 2000). Such current scales serve as guidance when creating validity scales for other measures.

Current Validity Scales Detecting Validity Threats

The MMPI was one of the first measures to create and emphasize the use of validity scales (Archer, Fontaine, & McCrae, 1998). This characteristic may be the source of psychometric strength that has resulted in the MMPI being the most widely used measure in forensic settings (Medoff, 1999). The MMPI, MMPI-2, and MMPI-A contain at least 7 different validity scales designed to identify a multitude of response biases. Some of these well-known scales include the Lie (L) and Defensiveness (K) scales which were both designed to detect underreporting of psychopathology in which a respondent attempts to present themselves in a favorable manner (Greene, 2000). The L scale measures deliberate dishonest responses with elevations in L resulting in lower elevation of most clinical scales of the MMPI. Content measured in the L scale includes denial of aggression and personal

dishonesties, bad thoughts, and small human failings. The K scale and K-correction measures defensiveness and identifies those individuals who present with significant psychopathology but whose profiles are within a normal limit. The MMPI's F scale, in contrast, identifies over-reporting of psychological symptoms and identifies individuals who answer test items in unusual or deviant ways. This scale contains items endorsed at a low rate in normative adults and adolescents to detect atypical responses and includes a wide range of content areas including peculiar experiences, contradictory or questionable beliefs, expectations, and descriptions of the self, strange thoughts, bizarre sensations, and feelings of isolation. The Variable Response Inconsistency (VRIN) and True Response Inconsistency (TRIN) scales are considered content free scale since both detect random responding rather than response accuracy. The VRIN scale examines items that have similar or opposite content to detect responses to pairs of items that are inconsistent. The TRIN scale detects inconsistent responses in same-paired items (Greene, 2000). Such validity scales for the MMPI have posed as a model for the creation of validity scales in other popular assessment tools (Briere, 1996; Greene, 2000; Morasco, Gfeller, & Elder, 2007; Morey, 1991).

Some other commonly-used measures that utilize validity scales are the NEO Personality Inventory Revised (NEO-PI-R), the Personality Assessment Inventory (PAI), the Millon Clinical Multiaxial Inventory II (MCMI-II), and the Traumatic Symptom Checklist for Children (TSCC). The NEO-PI-R includes validity scales to identify inconsistent responding, attempts by the respondent to present themselves in a positive light, and attempts to present themselves negatively (Morasco et al., 2007). Similarly, the PAI includes four validity scales to detect inconsistency in responses, infrequent responses, "faking bad" or over-reporting unlikely symptoms, and "faking good" or underreporting negative qualities

(Kurcharski et al., 2007). The MCMI-II includes scales to detect the extent to which respondents are over- and under-reporting characteristics, as well as a scale identifying respondents that failed to read or thoughtfully answer questions (Bagby, Toner, Gillis, & Goldberg, 1991). Finally, the Traumatic Symptom Checklist for Children is currently one of the only child report measures that include validity scales and consists of an Under-response and Hyper-response scale to identify under- and over-response to items assessing adjustment to trauma (Fricker & Smith, 2001). A reoccurring theme emerges when examining current tools with validity scales. That is, most include a scale to identify inconsistent responding and over- and under-reporting of symptoms or behaviors.

Existing measures may also be examined to determine possible ways of designing validity scales to detect common threats of measurement. At least two techniques have been created to detect inconsistent responses in measurement. One involves rational/intuitive identification of pairs of items with opposite or similar content and adding a raw score point to the validity scale when respondents provide inconsistent answers to responses. This method was used in the creation of the inconsistency scale of the PAI, and the VRIN and TRIN scales of the MMPI-A. The VRIN scale of the MMPI-A was created by identifying 67 item pairs that have either similar or opposite content. Examples of items on this scale that are similar in content are “138. I believe that I am being plotted against” and “99. Someone has it in for me”. When a respondent answers inconsistently for these items, for example “false” for item 138 and “true” for item 99, a raw scale score point is added to the VRIN scale. A cutoff score for the VRIN scale is then used to determine whether or not a respondents profile is valid. Recommendations cutoffs for the MMPI is a raw score of 12, which means those who inconsistently endorse 12 or more item pairs may not have a valid

MMPI-A profile (Green, 2000). A second, similar technique for creating inconsistency scales is by pairing items that are empirically derived. For example, the NEO-PI-R uses highly correlated items when selecting paired questions to include in the inconsistency scales instead of using intuition to identify item pairs (Schinka, Kinder, & Kremer, 1997). Instead of choosing item pairs on the basis of content, item pairs with $r > .40$ were considered for the inconsistency scale. To calculate a raw scale score of the inconsistency scale, the difference in responses for item pairs is calculated and then added across all 10 pairs making up the scale.

Techniques of both intuitive and empirical identification of items have also been used to create over- and under-response validity scales in measurement. Over-responding, or “faking bad” scales, are derived by identifying items that describe unlikely symptoms or items that are infrequently endorsed in the deviant direction by respondents. A raw scale point of one is added when respondents endorse items included on the scale. The Frequency (F) Scale of the MMPI was derived in such manner. Item selection of the F scale was determined through intuitive and empirical identification; F scale items that were believed to detect atypical ways of responding were selected. The final F scale was derived by identifying items endorsed in the deviant direction by 10 percent or less of a normative sample. When respondents report items in the deviant direction, a raw scale point is added to the F scale and T-scores are derived. T-scores above a predetermined cut-off are considered invalid.

Under-reporting or “faking good” scales identify items that represent common human failings or deviant behaviors that are commonly endorsed by the general population. When respondents fail to endorse common deviant behaviors or symptoms, a raw scale point is

added to this scale. The L scale of the MMPI is an example of a validity scale measuring underreporting. Items of the L scale were rationally selected by identifying common human failings and serve to detect individual's deliberate dishonest responding. Similar to the F-scale, deviant responses result in a raw scale point added to the scale, which are subsequently derived into T-scores. Development of over- and underreporting validity scales often involve rational or empirical identification of items and determining cut-off scores for each scale. Table 1 reviews common threats to validity, current instruments with validity scales, and procedures used to detect invalid responses.

Though discrepancy scales have not emerged as a common strategy for identifying inaccurate responses, techniques used in the development of validity scales for other measures such as the Personality Assessment Inventory may serve as a model for the development of this scale. The PAI's Defensiveness index was designed to detect defensive responding and underrating of faults. One of the scale's scoring criteria includes identification of individuals with large differences in pairs or specific scales or subscales. Two such subscales include the dominance and verbal aggression scales with differences in scale scores greater than 10T considered to be a discrepant response (Baity, Siefert, Chambers, & Blais, 2007). Although this scale is designed to detect underreporting, points are added to the scale when the difference between scores on scales exceeds a certain value (Baer & Wetter, 1997). This same concept may be applied to the creation of a discrepancy validity scale. Scale scores measuring different behaviors or people with significant discrepancies in scales may be useful for identifying inaccurate responding. For example, if a parent is reporting behaviors on two different siblings, scores for one child that are significantly different from scores for another child may reflect polarization or inaccurate

Table 1.

Existing Measures and Procedures for Validity Scale Development

Identified Threat to Validity	Instrument	Scale Name as Solution	Procedural Solution
<i>Random responding</i>	MMPI	VRIN	Scale consists of 67 item pairs with similar or different item content. Raw score point added to scale when responses to item pairs are inconsistent.
	MMPI	TRIN	Creation of scale includes 23 item pairs keyed in the same direction (true-true or false-false). Raw score points are added to TRIN scale if respondents inconsistently respond (false-true).
	NEO-PI-R	Inconsistency (INC)	Ten item pairs derived empirically using criteria that $r > .40$ for item pairs. Inconsistent responses on item pairs result in points added to the scale score.
	PAI	Inconsistency (INC)	Scale includes rationally identified same item pairs. Scores on the scale increase when respondents answer item pairs inconsistently, thus indicating invalid data.
	MCMII-II	Validity Index	Items rationally identified as likely endorsed at low rates such as "I was on the front cover of several magazines last year". Sum of deviant responses on items make up the validity index.
	BASC-2	Validity Index (V)	Raw score point added to scale with respondent endorses nonsensical items that have been rationally derived as so.
<i>Over-reporting of symptoms or problem behaviors</i>	MMPI	Frequency (F)	Scale consists of 60 items that were endorsed in less than 10% of a normative population in a deviant direction. One raw scale point added to the scale when an item on the scale is scored in the deviant direction.

(Table 1
continued)

	NEO-PI-R	Negative Presentation Management (NPM)	Scale derived through use of rational and empirical identification of items identified as reflecting unlikely faults. Points are added to the scale score when respondents endorse unlikely faults.
	PAI	Negative Impression Management (NIM)	Data considered invalid when raw scores on the scale are high due to endorsement of rationally identified items detecting unlikely symptoms in normal individuals.
	MCMII-II	Debasement Measure (DEB)	Detect endorsement of rationally identified items thought to represent individuals faking bad or over-reporting problems. More points added to the scale results in greater likelihood of an invalid profile.
	TSCC	Hyper-response	Consists of items empirically derived by inclusion of items with response of “almost all the time” in less than 5% of a normative sample.
	BASC-2	F Index (faking bad)	Items identified as infrequently endorsed in a normative population. Greater scale score results in indication of invalid profile with test examiners alerted to proceed with caution when interpreting results.
<i>Under-reporting of symptoms or problem behaviors</i>	MMPI	Lie (L)	Detect endorsement of deviant responses to items rationally identified as common human failings.
	MMPI	Defensiveness (K)	Use of empirically derived items to detect individuals who present significant levels of psychopathology but produce profiles in normal range.
	NEO-PI-R	Positive Presentation Management (PPM)	Scale consists of 10 rationally and empirically identified items having extreme mean scores to detect individuals denying common faults. When respondents endorse items in the unlikely direction points are added to the scale.

(Table 1
continued)

PAI	Positive Impression Management (PIM)	Items on scale rationally identified as denying minor faults when endorsed in keyed directions. Scale score increases when item are endorsed.
MCMII-II	Desirability Gauge Scale (DES)	Raw scale points added to scale when respondents endorse items that have been rationally identified to detect individuals faking good.
TSCC	Under- response	Detect endorsement of items identified as least likely to receive a response of “never” in a normative sample. Endorsement of items represent a denial of common symptoms resulting in greater under-response scores
BASC-2	L Index (faking good)	Identifies when respondents endorse items in keyed directions infrequently when compared to items endorsed in normative population.

reporting of one or both children. It is in this way the development of a discrepancy validity scale makes use of existing strategies to create validity scales in identifying known threats to validity.

Some identified threats to validity lack current solutions for detecting inaccurate responses and may require the identification of new strategies for developing required validity scales. Please see Table 2 for examples of such threats. One possible threat includes low reading level and its impact on response validity. A strategy for assessing insufficient reading level is not currently in place. A unique procedural solution may be to add items to an existing measure that assess the respondent’s reading comprehension. These listed types of threats to validity may depend on the nature of the specific assessment in question. For example, polarization on reports of behavior and low reading level as presented in Table 2

are both possible threats to validity for the previously mentioned Co-parenting Behavior Questionnaire. Other assessments that involve completing the measure under time pressure may suffer from threats of inaccuracy due to speediness of completing the test. Threats of these types require the use of innovative techniques for developing validity scales to address the threat.

Table 2.

Examples of Unresolved Threats to Validity

Identified Threat to Validity	Examples of threat to demand of specific tasks	Procedural Solution
Polarization on reports of behaviors	Parent from divorced family may alienate child from ex-spouse leading to child reports of more negative parenting behaviors for alienated parent.	Points added to a discrepancy scale when mothering and fathering scores are highly discrepant.
Low reading level or other language barrier	Anxiety about completing task, presence of a reading disorder, or questionnaire written in language that is not native to respondents are possible threats to inaccurate responding.	Creation of items with right and wrong responses to assess reading comprehension of respondents. Low reading scores may be indicative of invalid data on measure.
Inaccuracy due to time limits or speediness	Timed tests such as tasks in IQ tests or standardized tests may result in inaccurate responding due to time constraints or greater performance anxiety.	Use of computerized or timed testing to measure item level response times. Unexpectedly fast response times compared to a normative sample may indicate inaccurate responding on measure.

General Validity Scale Model

A general model in the construction of validity scales is proposed after reviewing the literature and scales currently used in measurement (Please see Table 3.). This model poses 3 major underlying processes including:

1. Identifying possible threats to validity,
2. Developing strategies to test the construct in question, and
3. Testing and norming the validity scales created.

Table 3.

Steps of the General Validity Scale Model

Steps		
1) Identify threats to validity	2) Identify procedures for detecting invalidity	3) Testing and Norming Validity Scales
<u>Subtype 1)</u> Threats Common to all measures with existing solutions (i.e. inconsistent/random responding, overreporting and underreporting behaviors of symptomology)	Strategies common to current measures	Existing or novel strategies
<u>Subtype 2)</u> Threats unique to the demands of the task or testing instrument, with existing solutions	Identify other measures with similar existing strategies	Existing or novel strategies
<u>Subtype 3)</u> Threats unique to the demands of the task or testing instrument, with no existing solutions	No currently existing strategy to address threat. Creation of unique procedure necessary.	Existing or novel strategies

Identifying threats to validity. At step one of the model, scale developers discover possible reasons for invalidity. Becoming familiar with the tasks and application of the measurement tool being developed is essential in recognizing threats to validity that may occur in assessment. Threats to validity may be one of three sub-types including: common threats with common solutions, unique threats with common solutions, and unique threats with unique solutions. Validity Subtype One involves recognizing possible threats common to many measurement tools with existing procedures for identifying invalid data. This includes the previously discussed threats of inconsistent or random responding and over- and under-reporting of symptoms or behaviors. These threats have existing procedures for identifying when validity is at risk and procedures for assessing presence of the validity threat.

Validity Subtype Two includes unique threats to the testing instrument for which existing strategies for detecting invalidity exist. These threats may be a product of the demands of the individual test or situational context of the evaluation and should also be recognized. The large numbers of items on unusually long questionnaires and the resultant time-demand for test completion may pose as a threat to validity when factors such as fatigue become an issue. This threat, unique to certain measures may have existing techniques such as the use of an inconsistency scales to detect invalidity. As another example, parental alienation may result in large discrepancies in child reported scales of mothering and fathering on the Co-parenting Behaviors Questionnaire. Solutions for this threat to validity may lie in the development of discrepancy scales which detect large differences in scale scores. Both of these examples evidence unique threats to validity that may result from the

nature of the instrument in question but can be addressed use already existing strategies to identify threats.

Validity Subtype Three are those unique threats for which strategic solutions for measuring the sources of invalidity are non-existent. Such threats may include low reading comprehension or test speediness that may interfere with accurate responding. The presence of this subtype of validity threat requires new procedures for identifying invalid responding. Addressing low reading comprehension may require the inclusion of items with correct and incorrect responses imbedded within a measure assessing the respondents reading level. When respondents do not endorse items correctly to reach a predetermined cutoff or reading level, the completed measure may be considered invalid. All unique threats fall under subtype three when existing strategies are not relevant to the threat in question. Whether the aforementioned misrepresentations in data for the three differing subtypes are intentional in nature is of less importance than detecting the likely threats to valid data.

Strategies for detecting invalid data. Developing procedures to assess possible threats to validity is the second process in the creation of validity scales under the General Validity Scale Model. Designing procedures to identify invalidity may include the use of common, existing techniques to measure threats to validity. Step Two may also require the development of new strategies to measure unique threats of validity. This will be the case when existing procedural strategies do not exist. Frequently used methods of identifying invalid data for common threats to validity are seen in existing measures with validity scales. Empirically or rationally identifying item pairs and adding a point to the validity scale when item pairs are answered in a conflicting manner may measure inconsistent responding. Creation of over- and under- responding scales may include identifying items that are not

commonly endorsed or are commonly endorsed in the general population and adding a raw score point to the respective validity scale when items are endorsed in the opposite direction expected. Scale scores greater than a pre-determined cut-off score may indicate invalid results of the measurement test. Procedures identifying threats to validity resulting from demands of the testing instrument will vary to meet the needs of the individual measure. Novel methodologies for creating validity scales may be necessary to address the specific threat in question. Examples of the subtypes of validity threats and procedures for creating validity scales are displayed in Table 4.

Table 4.

Threats to Validity, Threat Subtypes, and Procedural Solutions for Addressing Threats

Step 1: Identify Threat			Step 2: Procedural Solution	
Categorize Threat and Subtype	Identified Threat to Validity	Instrument	Scale Name as Solution	Procedural Solution
Validity Subtype One: Common Threat/ Common Solution	Random responding	MMPI	VRIN	Scale consists of 67 item pairs with similar or different item content. Raw score point added to scale when responses to item pairs are inconsistent.
	Random responding	NEO-PI-R	Inconsistency (INC)	Ten item pairs derived empirically using criteria that $r > .40$ for item pairs. Inconsistent responses on item pairs result in points added to the scale score.
	Over-reporting of symptoms or problem behaviors	MMPI	Frequency (F)	Scale consists of 60 items that were endorsed in less than 10% of a normative population in a deviant direction. One raw scale point added to the scale when an item on the scale is scored in the deviant direction.
	Over-reporting of symptoms or problem behaviors	NEO-PI-R	Negative Presentation Management (NPM)	Scale derived through use of rational and empirical identification of items identified as reflecting unlikely faults. Points are added to the scale score when respondents endorse unlikely faults.
	Under-reporting of symptoms or problem behaviors	MMPI	Lie (L)	Use of empirically derived items to detect individuals who present significant levels of psychopathology but produce profiles in normal range.
	Under-reporting of symptoms or problem behaviors	NEO-PI-R	Positive Presentation Management (PPM)	Scale consists of 10 rationally and empirically identified items having extreme mean scores to detect individuals denying common faults. When respondents endorse items in the unlikely direction points are added to the scale.

(Table 4 continued)

Validity Subtype Two: <i>Unique Threat/ Common Solution</i>	Long Administration Time	MMPI	Frequency and Back Frequency (F & F _B)	Identifies random responding on the second half of the test and compares to response patterns on the test as a whole to detect likely fatigue.
	Parental Alienation	CBQ	Discrepancy Scale	Differences in Mothering and Fathering scales that are statistically greater than discrepancies in normal discrepancies.
	Low Reading Level		Reading Comprehension	Creation of items assessing reading comprehension.
	Inaccuracy due to time limits or speediness		Speed	Computerized testing measuring item level response times.

Testing and norming validity scales. Testing and norming scales for detecting misrepresented data is the third step and an essential end piece in the successful development of validity scales. An instrument's validity scales should be tested with a population and setting for which the measure was intended in order to assure accuracy of scales. Several research methodologies have been used to test current validity scales including simulation design, differential prevalence design, and known-group designs (Baer & Miller, 2002). Simulation design is the most common way to test validity scales and it involves testing validity scale scores for a group instructed to intentionality bias their responses in a specified manner to a group completing the measure under standard instructions. The simulation design may be between groups in nature comparing two independent samples who receive different instruction sets for completing a measure. A study of the underreporting scale of the PAI is an example of a between subjects simulation design to test the usefulness of validity scales (Baer & Wetter, 1997). A group of undergraduate college students instructed to

underreport symptoms on the PAI had higher scores on the test's validity scales than a separate group given standard instructions for completing the measure. Within groups simulation design may also be used to compare two validity scores for one sample completing the measure once under standard instructions and again when instructed to feign in a way that would likely result in high validity scale scores (Baer & Miller, 2002).

Differential prevalence design is a second design to test validity scales by comparing validity scores for individuals who are believed to have strong motives for distorting their responses on a measure to scores for individuals without these incentives. For example, a random sample of undergraduates given standard instructions for a measure may be compared to scores for a group of individual's currently involved in a custody evaluation with an incentive to fake their answers. Known-group design is a third design used to analyze validity scale scores from a group who is known to feign to a group of individuals identified as responding truthfully. Though this design may have great external validity, one must use accurate methodology in determining who is feigning from those who are not. An example of this methodology may include comparing validity scale scores from a normative sample to validity scale scores from a group identified by clinicians, parents, or others as over-reporting symptoms or behaviors (Baer & Miller, 2002). The known-group design has been used to test the validity scales of the MMPI-2 to compare police applicants who admitted to misrepresenting themselves on the measure to applicants who were believed to answer truthfully (Borum & Stock, 1993). Each methodology has advantages and disadvantages; however, whichever method chosen, testing validity scales serves as useful step in the development of final scales.

Finally, norming the validity scales allow test developers to set cut-off scores in determining when a measure or parts of a measure are to be considered inaccurate. Depending on the nature and extent of invalid responding, the examiner may choose to discard test scores or retain test scores but interpret with caution. In other instances, examination of validity scales may provide insight into the psychological mechanisms interfering with inaccurate responding of the examinee. Using the three proposed steps of the General Validity Scale Model may show great promise as a guide for the development of validity scales for measures in which invalid responses are likely.

Application of the General Validity Scale Model to the CBQ

The Co-parenting Behavior Questionnaire measures both parenting and co-parenting behaviors in families of divorce. Children are asked to report on the co-parenting scales assessing cooperation and conflict as well as individual parenting behaviors for both the mother and father. Applications of the CBQ often involve assessing parenting in traditional clinical and forensic settings in which dysfunctional parenting and/or co-parenting are suspected. Its application may range from identifying maladaptive parenting strategies that aid in the treatment of families to using the measure in custody evaluations where potentially life-altering decisions are made. Whether intentional or not, misrepresentations of responses in child-reported ratings of parenting and co-parenting behaviors are not uncommon. When using the CBQ in custody evaluations, invalid scale scores may result in poor assignment of custody. Such an example evidences the necessity of validity scales when decisions that greatly shape the child's development and adjustment are in question.

Following the General Validity Scale Model, the successful creation of scales to assess the threats to validity may be accomplished for the Co-parenting Behavior

Questionnaire. Of first priority is identifying the threats to validity when completing the CBQ. Next, possible strategies to measure these threats can be considered. Finally CBQ validity scales should be tested and normed with 10 to 18 year olds for whom the measure is intended. The steps for creating validity scales of the CBQ using the General Validity Scale Model as a guideline is displayed below in Table 5.

Table 5.

Application of the General Validity Scale Model to the CBQ

Threats (Step 1)		Strategies (Step 2)	Norming and Testing (Step 3)
Validity Threat	Threat Subtype		
Negative Parenting Presentation	Common Threat/Common Solution (I)	Inclusion of items empirically identified as unlikely to be endorsed in deviant direction in a normative population. Items finalized by validating on a sample for which clinician's have rated children's likeliness to present parents negatively.	Compare raw scale scores for sample of children completing CBQ under standard instruction set to a sample instructed to present their parents negatively on the CBQ.
Positive Parenting Presentation	Common threat/Common Solution (I)	Detect endorsement of empirically and rationally identified items to identify individuals reporting denial of negative parenting for mother and father. Items included on scale are those answered in deviant direction that are endorsed in less than 15% of the normative population. Validate on a clinical sample of children using clinician ratings of likeliness to present parents positively.	Compare raw scale scores for sample of children completing CBQ under standard instruction set to a sample instructed to present their parents positively on the CBQ.
Inconsistent Responding	Common Threat/Common Solution (I)	Inconsistencies in same or different item pairs rationally and empirically derived using a normative sample of children. Determine finalized items on scale by examining items likely to predict clinician's ratings of a child's inconsistent responding.	Compare scale scores for a normative sample to scale scores determined from computer generated responses to the CBQ.

(Table 5
continued)

Parental Polarization	Unique Threat/Common Solution (II)	Create item pairs by pairing each mothering item with each corresponding fathering item that is equal in content. Determine a discrepancy score by subtracting scores from the mothering and fathering scale. Determine final items by examining those items which best predict likeliness for the child to rate parents in a highly discrepant manner as identified by the clinician.	Compare raw scale scores for sample of children completing CBQ under standard instruction set to a sample instructed to present their parents in a highly discrepant manner (either high on mothering and low on fathering or low on mothering and high on fathering).
Low Reading Level	Unique Threat/Unique Solution (III)	Inclusion of items to assess appropriate reading levels for 10-18 year olds. Validate items on sample identified by clinicians as likely to experience reading difficulties.	Compare reading comprehension validity scale scores from a normative sample randomly generated responses on the CBQ.

Step 1. Identifying threats to validity. Following the first step of the General Validity Scale Model, identification of the threats of validity of the Co-parenting Behavior Questionnaire is determined. The three aforementioned threats to validity that are common to all measurements may be applied in measuring parenting and co-parenting behaviors from the child’s perspective. These threats include random or inconsistent responding, over-reporting negative or under-reporting positive parenting and co-parenting behaviors, and under-reporting negative or over-reporting positive behaviors. Random responding may occur for several reasons such as the child wanting to stay as far out of the divorce conflict as possible or wanting to avoid pressure of choosing sides between parents. Inconsistent responding may result from inattention to completing the task at hand. Children may also be likely to over and underreport mothering, fathering, and co-parenting behaviors whether intentional or not. Under-reporting positive or over-reporting negative parenting and co-parenting behaviors, which can be referred to as negative parenting presentation, may occur for reasons such as the child feeling resentment toward one or both parents for divorcing.

Children may be prone to a positive parenting presentation when over-reporting positive or under-reporting negative parenting and co-parenting behaviors if they are concerned about hurting their parents' feelings, are overprotective of their parents, feel caught between parents, or perhaps even fear one or both parents. These three common threats to validity have already existing solutions for other measures and are thus categorized as Subtype I threats.

Within the CBQ there exist two unusual threats to validity. Parental polarization, or experiencing an exaggerated or a contrived distance from one parent is an unfortunate process unique to high conflict divorces. The polarization process will distort a child's assessment of one or both parents. The CBQ includes scales of both mothering and fathering parenting behaviors and as such this may result in excessively discrepant responses of the parents. That is, the child reports one parent in an unusually positive light while reporting the other parent in an excessively bad light. Reasons for high polarization of responses may include that the child experiences elevated levels of triangulation of parents, rules may not be as strict at one parent's house, or the child may feel one parent in particular needs them. Developing procedures for identifying such a threat to validity may involve examining current validity scales of the Personality Assessment Inventory. As a unique threat to validity with a possible suggested solution, parental polarization may be categorized as a Subtype II threat.

Low reading comprehension may pose as a unique threat to CBQ validity with necessary unique solutions thus categorized as a Subtype III threat. The Co-parenting Behavior Questionnaire is intended for use for ages 10-18, and accurate responding may require that respondents have an adequate reading level for comprehending each item. Low

reading comprehension may result in random responses that are unintentional or unrelated to fatigue. Assessing low reading comprehension may require the creation of new procedural solutions to test whether children have the necessary reading level to complete the CBQ.

Step 2. Strategies for detecting invalid responding. Determining relevant procedures for detecting validity threats of the Co-parenting Behavior Questionnaire is the next process in validity scale development when following the General Validity Scale Model. Goals in this step are twofold. First, an initial item pool for each validity scale will be created. For most scales, this will be accomplished using a normative sample of children from divorced households to determine relevant items for each scale. Secondly, the pool of items identified for each scale will be validated and tested with a clinical sample to determine final item inclusion. Developing an initial item pool for each scale may be guided by existing validity scales of other measures.

The validity scales proposed to identify children's inaccurate responding on the CBQ include a negative parenting presentation scale, a positive parenting presentation scale, an inconsistency scale, a reading comprehension scale, and a parenting discrepancy scale. The Negative Parenting Presentation (NPP) Scales will include three validity subscales to identify when children give overly negative reports of invalidity in reports of each mothering, fathering, and co-parenting on the CBQ. Positive Parenting Presentation Scales (PPP) will also include three subscales for mothering, fathering, and co-parenting and will be used to detect overly positive reports of parenting and co-parenting. Development of an inconsistency scale will be helpful in detecting children's inconsistent or random responding on the CBQ. These first three validity scales are designed to detect common validity threats and can be developed using existing common solutions. A new strategy to detect random

responding on the CBQ due to the presence of a low reading level is required since current solutions do not exist. Lastly, a Parenting Discrepancy or polarization scale may be one strategy to assess the process of polarization in which children favor one parent and thus have distorted responses on the CBQ, even if both parents are competent. As seen with the development of a discrepancy scale and other proposed scales of the CBQ, the General Validity Scale Model may serve as guide in the creation of validity scales for any given measure. Designing each validity scale may use existing strategies or may involve new unique strategies to identify threats of validity. The purpose of Study 1 is to use new and existing strategies to develop validity scales to detect each of the 5 proposed validity threats of the CBQ. This study focuses specifically on item inclusion for each of the 5 validity scales. The procedures for the creation of each validity scale are outlined in detail in the methodology section.

Step 3. norming and testing validity scales. The third step of the General Validity Scale Model involves norming and testing validity scales to assure that the validity scales are themselves valid. To test validity scales of the CBQ, the simulation design may be used by comparing a sample of children receiving standard CBQ instructions to a sample instructed to respond to the CBQ in a biased manner. These comparisons may involve between or within group comparisons to compare validity scores for the two instruction sets for each validity scale. It is expected that the validity scale scores under the standard instruction set will be significantly lower than the validity scale scores under the biased instruction sets.

The purpose of Study 2 is to test each validity scale created in Study 1. Future studies should focus on norming the scales and creating scoring criteria for each validity scale. Specifically in Study 2, a within groups simulation design will be used to test the potential

NPP, PPP, an Discrepancy scales developed in Study 1 by comparing raw scale validity scores given to a sample completing the CBQ under standard instructions to scores from the sample that were simulated in some way. The Inconsistency and Reading Comprehension scales will be tested by comparing validity raw scale scores from a sample given standard instructions to raw scale scores determined from computer generated random response sets on the CBQ. The methodology section further outlines the procedures used to test each validity scale in Study 2.

Hypotheses

The General Validity Scale Model serves great utility in guiding the development of validity scales for any measure. Such use is exemplified in the planned development of several validity scales for the Co-parenting Behavior Questionnaire. Given its use in making important life decisions for families of divorce, there is a great need for CBQ validity scales. The current study is designed to apply the general scale model through the item development and validation of validity scales for use with the measure.

1. It is hypothesized that there does exist certain threats to children's accurate rating of their parents' parenting and co-parenting behaviors and that these threats to valid reporting can be identified and measured. These anticipated threats to validity include: 1) Inaccurate responding due to overly negative presentations of parenting and co-parenting 2) Inaccurate responding due to overly positive reporting of parenting and co-parenting 3) Inaccurate responding due to polarized or extreme bifurcation in ratings of each of their parent's parenting and co-parenting behaviors, 4) Inaccurate responding due to inconsistent or random responding, and 5) Inaccurate responding due to low reading comprehension.

2. It is further proposed that five different validity scales may be developed to accurately detect each of these threats to the CBQ. A Negative Parenting Presentation scale will be developed to detect overly negative reports of parenting and co-parenting by determining and including items answered in a deviant direction than that of a normative population. Similarly, a Positive Parenting Presentation scale can be used to identify children giving overly positive reports of parenting behaviors by including items answered in the unexpected direction in less than 15% of a normative population onto the scale. Items pairs that assess largely discrepant responses in parental behaviors will be included on a Discrepancy scale to detect inaccurate responding due to parental polarization. An inconsistent or random responding scale will use empirically and rationally determined same and different item pairs to identify children who respond inconsistently or randomly on the CBQ. Lastly, a reading comprehension scale will be developed by creating new items assessing the child's ability to understand current items on the CBQ.

3. The five validity scales will be tested in Study 2. It is hypothesized that the validity scales created in Study 1 may accurately detect intentionally distorted response sets from response sets completed under standard directions. Specifically, The Negative Parenting Presentation, Positive Parenting Presentation, and Discrepancy scales will be tested by comparing raw validity scale scores determined under a standard instruction set to the same sample instructed to respond to the CBQ in a biased manner thus elevating these validity scale scores. It is hypothesized that the NPP, PPP, and Parenting Discrepancy scale scores will be significantly lower for the standard instruction set than they will be when instructed to feign. The

Inconsistency scale and Reading Comprehension scale will compare raw scale validity scores under a standard instruction set to a set of randomly generated item responses that are believed to result in high scale scores for these two validity scales. For the Inconsistency scale, it is hypothesized that raw scale scores under the standard instruction set will be significantly lower than scale scores computed from a set of randomly generated responses. For the Reading Comprehension scale, it is believed that raw scale scores will be greater when participants completed the CBQ under standard instructions than for raw scale scores computed from a set of randomly generated item responses.

Methods - Study 1

Participants

The sample for Study 1 consists of a normative sample of 517 children whose biological parents are divorced and one parent of each child. Data from this sample were used in a previous study evaluating the Co-parenting Behavior Questionnaire (Schum, 2003). The participants include children, younger siblings, or acquaintances of undergraduate psychology majors at a public university in Virginia. Undergraduates who identified participants or participated in the study did so to complete a course requirement. The child participants' ages ranged from 10 to 18 years old. The sample consisted of 319 Caucasians, 147 African-Americans, 11 Hispanics, 6 Middle Easterners, 4 Asians, and 30 Multi-racial children. Parents reported 74% of mothers having sole physical custody, 10% of fathers having sole physical custody, and 15% of joint physical custody. Reports of legal custody involved 53% mother, 7% father, and 40% joint custody.

Measures

Co-parenting Behavior Questionnaire (CBQ). The CBQ is a child-report measure of parenting and co-parenting behaviors in post-divorce families (Macie & Stolberg, 2003; Schum & Stolberg, 2007; Stolberg, Ferrante & Schum, 2006). Items are reported on a 5-point Likert scale ranging from 1) *almost never* to 5) *almost always*. The CBQ consists of 12 subscales including 4 co-parenting dimensions (Parental Conflict, Triangulation, Co-parental Communication, Co-parental Cooperation), as well as 4 mothering and 4 fathering behaviors measuring Warmth, Monitoring, Parent-Child Communication, and Discipline. The CBQ demonstrates good internal consistency with an alpha coefficient of .93. The original CBQ was a 92-item measure that was later revised into a shortened 86-item tool. Data from Study 1 was collected using the 92-item version of the CBQ with extra items deleted from the dataset to reflect the 86 item version. *Appendix A* shows which item numbers of the CBQ load onto each of the parenting and co-parenting scales.

Demographics Questionnaire – Parent Report (DQ). The DQ is a parent report of standard demographic information including 17 items to assess for age, gender, and ethnicity of the child and parent as well as questions related to parental divorce.

Procedures

Child participants from the sample were given ID numbers and provided with packets including detailed instructions, a consent form, and the Co-parenting Behavior Questionnaire. One parent of the child signed the consent form agreeing for the parent and child to participate in the study and completed the Demographics Questionnaire. Each child completed the CBQ. Data from this normative sample of children from divorced families was used in attempts to create an initial item pool for Negative Parenting Presentation, Positive

Parenting Presentation, Inconsistency, and Discrepancy scales of the CBQ. Development of validity scales in this study focused on determining item inclusion on each scale, while study 2 focused on validating each scale. Strategies in the development of each validity scale are described below.

Scale construction of the Negative Parenting Presentation scales. Three possible NPP validity scales were identified to detect children who report overly negative mothering, fathering, and co-parenting. Creating an initial item pool for the Negative Parenting Presentation (NPP) scales utilized strategies similar to those currently employed in measures such as the MMPI or Traumatic Symptom Checklist for Children. Development of the NPP scales involved using the normally distributed data set of the normative sample to empirically identify items on the parenting and co-parenting subscales that are endorsed in a deviant direction in less than 15% of the population. Identifying items endorsed in the unexpected direction from that of the normative sample was used to detect children reporting overly negative parenting behaviors.

Items describing negative parenting and co-parenting behaviors that are endorsed as almost always-5 and positive behaviors endorsed as almost never-1 in less than 15% of the normative sample were included on the appropriate subscale. This implies that when children answer in an unexpected direction and magnitude (unusually negative) on the NPP scales, they are giving an overly negative response pattern that is not typical in the normative population of children. Items meeting criteria that were identified as representing mothering were included on the Negative Parenting Presentation Mothering Scale, while those items describing fathering and co-parenting were included on the NPP-Fathering Scale and NPP-Co-parenting Scale, respectively.

Scale construction of the Positive Parenting Presentation scales. Potential items on the Positive Parenting Presentation (PPP) scales were determined in much the same fashion. Three possible subscales for mothering, fathering, and co-parenting were identified to be used to detect overly positive reports on the CBQ. The normative sample was used to identify items endorsed by few individuals of the population in a direction implying over-reporting positive or under-reporting negative parenting and co-parenting. Items to be included in this item pool were those describing positive parenting behaviors that were endorsed as almost always-5 or negative parenting behaviors endorsed as almost never-5 by 15% or less of the sample. The resulting items were examined for content to determine inclusion on a PPP-Mothering, PPP-Fathering, and PPP-Co-parenting scales.

Scale construction for the Parenting Discrepancy scale. Though a threat unique to the CBQ, the development of a parenting discrepancy scale followed strategies used in existing scales such as the Personality Assessment Inventory, in which large differences in scales may indicate invalidity. First, an initial item pool was determined by creating item-pairs from the mothering and fathering scales. Each of the 25 mothering items on the CBQ has a corresponding fathering item with similar content resulting in an initial 25 item pairs for the scale. Once these items were identified, a discrepancy score for each item pair was determined. This was done by subtracting each mothering score from each fathering score for every single item and taking the absolute value of this score. Using similar techniques for the over- and under-reporting scales, item pairs were examined to identify pairs in which extreme discrepancy scores occur in less than 15% of the population. This identified item pairs that are not likely to be answered in a highly discrepant way in the normative population. These item pairs were retained on the discrepancy scale.

Scale construction for the Inconsistency scale. To detect inconsistent and random responding on the CBQ, the normative sample and current strategies used in other measures in developing inconsistency scales were employed. CBQ items were empirically and rationally identified to create same and opposite item pairings. First, item pairs were empirically identified with those item pairings having $r > .40$ being considered for the final scale. A correlation of $.40$ has been used in previous development of an Inconsistency Validity scale of the NEO-PI-R, and was chosen because it was believed these items would demonstrate sufficient empirical support to be similar enough in content (Morasco et al., 2007). Item pairs with $r > .40$ were examined for content to confirm that correlations are in the desired direction and make intuitive sense. When children answer in substantially different ways to same-item pairings or similarly to opposite-item pairings on the scale, it will be interpreted to indicate random or inconsistent style of responding. For example, after identifying an item pair with a correlation of $.40$ or larger, if the item “I talk to my mom about my problems” is endorsed as almost always, but the same-paired item “If I have a problem, I talk to my mom about it” is endorsed as almost never then responses are deemed inconsistent. Items pairs that meet statistical and rational criteria were retained on the CBQ inconsistency scale.

Scale construction for the Reading Comprehension scale. Random responding on the CBQ due to the presence of a low reading level may be detected through the development of a unique validity scale for which existing scales do not exist for other measures. The Reading Comprehension scale required the addition of new items for the CBQ specifically designed to ensure that the child respondent is capable of reading and interpreting existing items on the measure. Seven new items were developed containing the same wording as 7

current CBQ items describing parenting and co-parenting from divorced households. The new 7 items asks children to identify the answer that best describes the meaning of the item from a list of five possible responses. Two reading comprehension items from the scale describe mothering behaviors, two describe fathering, and three describe co-parenting behaviors. The reading comprehension items were inserted at the end of the CBQ as not to influence the child-respondent's interpretation and response to the CBQ content items. The order in which the reading comprehension items are presented is from those of a lower reading level to those of a more difficult reading level. The reading level used for ordering items was determined using the Flesch-Kincaid Readability Test indicating at which grade level each item is.

Results – Study 1

Study 1 was conducted to test the hypotheses that five different validity scales may be developed to accurately detect validity threats of the Co-parenting Behavior Questionnaire. Previously collected data from the 86 items of the CBQ were used to test for and determine item inclusion on the Negative Parenting Presentation, Positive Parenting Presentation, Parenting Discrepancy, and Inconsistency Scales. Empirical and rational identification of items was employed for development of each of these scales. Development of the Negative Parenting Presentation Scale and Positive Parenting Presentation scale were attempted by examining item level frequencies. Items in which 15 percent or less of the population answered in a deviant direction indicating overly negative or overly positive parenting were identified and obtained on their respective scales. For the Discrepancy scale, item pairs were rationally identified by pairing mothering items with similarly worded fathering items. After calculating a discrepancy score between item pairs, frequencies were run to determine item

pairs in which 15% or less of the population had a large discrepancy score (score of 4). Item pairs for the inconsistency scale were identified by those pairs with $r > .40$ with item pairs making intuitive sense were retained on the final scale. The Reading Comprehension Scale was constructed through the addition of 7 new CBQ items and considering the reading level of each item. A more detailed description of the statistical analyses involved in creating the validity scales is provided below.

To test the hypothesis that a Negative Parenting Presentation scale could be developed to detect overly negative reports of parenting and co-parenting behaviors frequencies of item-level responses were run for the 86 CBQ items. Item frequencies in which less than 15% of the population answered CBQ items in a deviant (overly negative) direction were identified. Items endorsed as 1-almost never and 5-almost in an overly negative way by less than 15% of the population were retained for the Negative Parenting Presentation scale. Table 6 displays items that met this criteria as well as item level responses (1 or 5), and item level frequencies. Items were also examined for content to determine whether they loaded onto the mothering, fathering, or co-parenting scales on the CBQ. The subscale each item loads onto is also indicated in Table 6.

Twenty-three items met initial criteria for inclusion in the Negative Parenting Presentation Co-parenting subscale. A total of eight finalized items were included on the NPP co-parenting scale after additional inclusion criteria were applied. To reduce final item inclusion on the scale to 8, items that indicated specific behaviors of the mother toward the father or father toward the mother were deleted. For example, items such as “7. My mom

Table 6.

CBQ Items, Frequencies, and Subscale Loadings for NPP Scales

Item Number	Item Level Response (1 or 5)	Item Level Frequency	Subscale Loading
1	5	13.7	Co-parenting
2	5	7.9	Co-parenting
3	5	8.9	Co-parenting
4	5	10.2	Co-parenting
5	5	10.8	Co-parenting
7	5	8.1	Co-parenting
8	5	9.1	Co-parenting
9	5	7.7	Co-parenting
10	5	3.9	Co-parenting
11	5	7.1	Co-parenting
12	1	13.9	Co-parenting
13	5	11.8	Co-parenting
14	5	9.1	Co-parenting
15	5	7.3	Co-parenting
18	1	11.8	Co-parenting
19	5	5.2	Co-parenting
21	5	11.4	Co-parenting
22	5	11.8	Co-parenting
25	5	9.5	Co-parenting
27	5	6.9	Co-parenting
31	5	13.5	Co-parenting
33	1	11.2	Co-parenting
35	5	9.5	Co-parenting
37	1	6.4	Fathering
38	1	4.1	Mothering
39	1	7.1	Mothering
40	1	5.0	Mothering
41	1	3.5	Mothering
43	1	7.5	Fathering
44	1	8.9	Mothering
45	1	6.2	Mothering
46	1	6.9	Mothering
47	1	2.7	Mothering
48	1	4.4	Mothering
50	1	14.3	Fathering
53	1	8.5	Mothering
57	1	7.5	Mothering
58	1	5.6	Mothering
59	1	4.4	Mothering

(Table 6 continued)

62	1	14.5	Fathering
63	1	11.0	Mothering
64	1	12.2	Mothering
65	1	2.5	Mothering
71	1	8.5	Fathering
72	1	4.4	Mothering
73	1	14.3	Fathering
75	1	6.8	Mothering
78	1	7.7	Mothering
79	1	13.9	Fathering
80	1	4.4	Mothering
81	1	3.1	Mothering
82	1	4.1	Mothering
84	1	3.9	Mothering
86	1	8.3	Mothering

asks me questions about my dad that I wish she would not ask” and “9. My dad asks me to carry messages to my mom”. This was done to clearly distinguish co-parenting behaviors and items from mothering and fathering behaviors and items. Thus, if children have highly discrepant responses for mothers and fathers it should not impact their NPP Co-parenting score. Items that met criteria for the NPP Co-parenting subscale that used general language to speak about both parents were included on the final NPP scale. Examples of such items include “3. My parents argue about money in front of me” and “10. My parents fight about where I should live”.

In developing NPP fathering and mothering scales, twenty-four mothering items and seven fathering items met criteria for items endorsed in the deviant direction in less than 15% of the overall population. A total of 7 fathering and 7 mothering items were identified that met criteria and comprised final item inclusion for the NPP Fathering and NPP Mothering subscales. Items were examined for content to determine this final item inclusion for the separate NPP Mothering and NPP Fathering scales. Fathering items that met criteria for the

NPP Fathering subscale were paired with similarly worded mothering items that met criteria for the NPP mothering subscale. This was done to make the NPP Mothering and NPP Fathering subscales similar in nature to control for the parent in question.

Development of the NPP Co-parenting, Mothering, and Fathering scales support the hypothesis that a scale containing CBQ items can be developed to assess reporting overly negative parenting and co-parenting behaviors. Table 7 contains a list of finalized items that are included on the NPP Co-parenting, Mothering, and Fathering scales. Items retained on the NPP fathering and mothering scale include, “My mom (dad) and I have friendly talks” and “I feel that my dad (mom) cares about me”.

Table 7.

Finalized Item Inclusion for NPP Scales

Subscale Name	Items
NPP Co-Parenting Scale	<ul style="list-style-type: none"> <li data-bbox="846 1066 1377 1102">1. My parents complain about each other. <li data-bbox="846 1142 1424 1213">3. My parents argue about money in front of me. <li data-bbox="846 1253 1390 1325">4. When my parents argue, I feel forced to choose sides. <li data-bbox="846 1365 1406 1436">5. When my parents talk to each other, they accuse each other of bad things. <li data-bbox="846 1476 1312 1512">8. I feel caught between my parents. <li data-bbox="846 1551 1386 1623">10. My parents fight about where I should live. <li data-bbox="846 1663 1312 1698">14. My parents argue in front of me. <li data-bbox="846 1738 1424 1799">22. When my parents talk to each other, they get angry.

(Table 7 continued)

NPP Mothering Scale

- 38. My mom and I have friendly talks.
- 41. I feel that my mom cares about me.
- 47. My mom likes being with me.
- 65. My mom says nice things about me.
- 81. My mom praises me when I do something good at home or at school.
- 82. My mom says she loves me and gives me hugs.
- 84. My mom is patient with me.

NPP Fathering Scale

- 37. My dad likes being with me.
 - 43. I feel that my dad cares about me
 - 50. My dad says he loves me and gives me hugs.
 - 62. My dad praises me when I do something good at home or at school.
 - 71. My dad says nice things about me.
 - 73. My dad and I have friendly talks.
 - 79. My dad is patient with me.
-

The hypothesis that a Positive Parenting Presentation scale could be constructed to detect overly positive reports of parenting and co-parenting behaviors was tested in a similar fashion as the NPP scales. Item level responses were run to determine items in which less than 15% of the population answered CBQ items in a deviant, overly positive manner. Items answered as a 1-never or 5-always indicating positive parenting and co-parenting were

retained for the PPP scales. Examination of item content allowed for determination of loadings onto the co-parenting, mothering, or fathering PPP subscales. Table 8 displays items that met criteria for inclusion on the PPP scales, the item level response (1 or 5) for each item, item level frequencies, and subscale loadings. Examples of items meeting criteria include: “19. My parents talk to each other about how I feel about the divorce.” and “70. I talk to my dad about my problems.”

Table 8.

CBQ Items, Frequencies, and Subscale Loadings for PPP Scales

Item	Item Level Response (1 or 5)	Item Level Frequency	Subscale Loading
19. My parents talk to each other about how I feel about the divorce.	5	5.2	Co-parenting
34. My parents get along well.	5	14.5	Co-parenting
49. I have chores to do at my dad’s house.	5	13.5	Fathering
70. I talk to my dad about my problems.	5	14.3	Fathering
77. My dad talks to me about my friends.	5	12.2	Fathering
83. If I get in trouble at school, my father punishes me.	5	11.6	Fathering

Two co-parenting and four fathering items met criteria for the PPP Co-parenting and PPP Fathering scales respectively. No mothering items met criteria for the PPP Mothering scales. There was failure to construct a validity PPP scale due to the low number of items meeting criteria on any of the three subscales. Thus, the hypothesis that a validity scale could

be developed in the described manner to detect overly positive reports of parenting and co-parenting behaviors on the CBQ was not supported.

Testing the hypothesis that a Parenting Discrepancy scale could be developed to detect highly discrepant evaluations of mothering and fathering was accomplished in three steps. First, item-pairs were created consisting of 25 mothering items matched with 25 similarly worded fathering items on the CBQ. Second, discrepancy scores for each item pair were calculated by subtracting numbered responses on the mothering item from corresponding fathering item responses and taking the absolute value of the difference scores. Third, frequencies were determined for each of the 25 discrepancy scores. Extreme discrepancy scores (score of 4) for item pairs that occurred in less than 15% of the population were identified and retained for an initial item pool for the discrepancy scale.

Eighteen out of the 25 item pairs met initial criteria for inclusion in the Parenting Discrepancy scale. Two item pairs from each of the four different CBQ mothering and fathering subscales were chosen to construct final item inclusion on the Parenting Discrepancy scale that tapped the breadth of important parenting behaviors assessed in the CBQ. Thus, two item pairs were chosen from the monitoring subscale, two from the communication subscale and two each from the warmth and discipline subscales. The two item pairs chosen from each subscale were item pairs with the lowest frequency for discrepancy scores. A total of 8 item-pairs were retained for the final Parenting Discrepancy scale supporting the hypothesis that a validity scale could be developed in attempts to detect inaccurate reporting due to parental polarization. Item numbers, frequencies, and subscale names for each pair are displayed in Table 9. Examples of included item pairs retained on the Discrepancy scale are “65. My mom says nice things about me.” paired with “71. My dad

says nice things about me.” and “80. I talk to my mom about things that I do well.” paired with “85. I talk to my dad about things I do well.”

Table 9.

Item Pairs, Frequencies, and Subscale Loadings for the Parenting Discrepancy Scale

Item Numbers	Frequency of Extreme Discrepant Responses	Parenting Subscale
60. If I have problems in school, my mom knows about it. 52. If I have problems in school, my dad knows about it.	13.5%	Monitoring
65. My mom says nice things about me. 71. My dad says nice things about me.	5.2%	Monitoring
80. I talk to my mom about things that I do well. 85. I talk to my dad about things I do well.	7.1%	Communication
38. My mom and I have friendly talks. 73. My dad and I have friendly talks.	5.6%	Communication
41. I feel that my mom cares about me. 43. I feel that my dad cares about me.	6.0%	Warmth
47. My mom likes being with me. 37. My dad likes being with me.	3.9%	Warmth
84. My mom is patient with me. 79. My dad is patient with me.	7.7%	Discipline
86. When my mom says she is going to punish me, she does it. 74. When my dad says he is going to punish me, he does it.	9.7%	Discipline

Item pairs were also constructed to test the hypothesis that an Inconsistency scale can be developed to identify inconsistent or random responding on the CBQ. The final inconsistency scale consists of 7 same item pairs and 3 opposite item pairs for a total of 10

item pairs. To determine item inclusion, same and opposite item pairs were empirically and rationally identified to determine item content on the Inconsistency scale. Empirical identification of items was accomplished by running item level correlations for all CBQ items. Item pairs in which Pearson's $r > .40$ were retained on the initial item pool for the inconsistency scale. Items were then examined for item content to rationally identify same and opposite item pairs. Items that met criteria were used no more than once for item pairs on the scale requiring deletion of some qualifying item pairs. For example, item 14 "My parents argue in front of me" was empirically and rationally identified as a possible pairing with items 34 "My parents get along well" and 35 "My parents yell at each other"; however, only one item pair (items 14 & 35) was retained on the final inconsistency scale. Item numbers, Pearson's r , and descriptions of pairings are presented in Table 10. Development of same and opposite item pairs supports the hypothesis that a validity scale may be developed for the CBQ to detect inconsistent and random responding.

Table 10.

Item Pairs for Inconsistency Scale

Item Pairs	Pearson's r	Item Pair Description
14. My parents argue in front of me. 35. My parents yell at each other.	0.661	Same item pair
29. My parents talk to each other about the good things I do. 23. My parents talk to each other about big choices in my life.	0.659	Same item pair
71. My dad says nice things about me. &62. My dad praises me when I do something good at home or at school.	0.705	Same item pair
55. My dad knows who my friends are and what they are like. 77. My dad talks to me about my friends.	0.680	Same item pair
75. My mom knows who my teachers are and how well I am doing in school. 60. If I have problems in school, my mom knows about it.	0.567	Same item pair
53. When I break one of my mom's rules, she punishes me. 64. If I get in trouble at school, my mom punishes me.	0.660	Same item pair
4. When my parents argue, I feel forced to choose sides 8. I feel caught between my parents.	0.670	Same item pair
5. When my parents talk to each other, they accuse each other of bad things. 6. My parents talk nicely to each other.	-0.520	Opposite item pair
22. When my parents talk to each other, they get angry. 34. My parents get along well.	-0.574	Opposite item pair
28. My mom tells me good things about my dad. 25. My mom tells me bad things about my dad.	-0.402	Opposite item pair

In order to test the validity of a reading comprehension scale in Study 2, such a scale was developed in this study. To create the scale, 7 new items were added to the CBQ as described in the methods section. To determine the order of presentation of the 7 additional items, the items were screened for grade level using the Flesch-Kincaid system. Table 11 presents the Reading Comprehension scale items and the associated Flesch-Kincaid grade for each. An example of an item developed for this scale is “89. *“When I break one of my dad’s rules, he punishes me.”* means that”.

Table 11.

Reading Comprehension Items and Grade Level

Reading Comprehension Scale Item	Flesch-Kincaid Grade Level
87. <i>“My mom and I have friendly talks.”</i> means:	0.6
88. <i>“When I leave the house, my dad knows where I am and who I am with.”</i> means that:	2.4
89. <i>“When I break one of my dad’s rules, he punishes me.”</i> means that:	2.6
90. <i>“I feel caught between parents.”</i> means that:	2.8
91. <i>“When I do something wrong, my mom talks to me about it.”</i> means that:	3.8
92. <i>“My parents talk to each other about big choices in my life.”</i> means that:	3.8
93. <i>“My parents complain about each other.”</i> means that:	6.4

Discussion - Study 1

The present study progressed the development of validity scales for the Co-Parenting Behavior Questionnaire. Overall, results from Study 1 support the development of validity scales for the Co-parenting Behavior Questionnaire using the first two steps of the General Validity Scale (GVS) Model. Five possible threats to validity were identified and strategies were developed to detect the specific threat in question. The development of these scales was unique in several ways including: 1) using the GVS Model as guidance for its creation and 2) using both empirical and rational identification of items and items pairs. This differs from previous development of validity scales that rely on a single technique, either empirical identification or rational identification, to determine item inclusion.

The Negative Parenting Presentation scale, Parenting Discrepancy scale, Inconsistency scale, and Reading Comprehension scale were developed using empirical and rational identification techniques. The creation of these scales supports the hypothesis that validity scales may be developed for the CBQ to detect overly negative reports of parenting and co-parenting, parental polarization, inconsistent responding, and low reading comprehension. The development of each validity scale also revealed patterns in typical CBQ responses that may affect validity scale scores or help an examiner make sense of CBQ responses. A discussion of each validity scale including support for hypotheses, the use of empirical and rational identification in the approach to development, and ways the development highlight the nature of the CBQ is provided in more detail below.

Results supported the development of three different NPP subscales including the mothering and fathering subscales with 7 items each and a co-parenting subscale with 8 total items. Empirical identification of items was important in determining an initial item pool for

the NPP scales while rational identification of items was useful in determining final item content on each of the three NPP scales.

Unique qualities of the CBQ were also revealed from the creation of the NPP scales. Specifically, it was found that empirical criteria for the NPP mothering scale resulted in an original 24 item mothering scale and only 7 items on the fathering scale. The infrequent endorsement of negative extreme scores for mothers' parenting suggests that children are over-reporting negative mothering behaviors less frequently than fathering behaviors on the CBQ. There may be several causes for this differential response pattern. For example, children may share more time with their mother than their father due to custody arrangements that lead to slightly skewed responses in reports of negative parenting. This finding highlights the discrepant nature that may potentially present in the responses for items loading onto the NPP fathering and mothering scales. Despite these response differences, Study 1 revealed a sufficient number of items loading onto the three NPP scales. This supports the development and usefulness of this validity scale to determine when children are providing an overly negative presentation of parenting and co-parenting.

Results were also consistent with hypotheses supporting the development of a Parenting Discrepancy scale with findings further shedding light onto the nature of the CBQ. Eighteen item pairs met statistical criteria for inclusion on the Parenting Discrepancy Scale. This infrequent reporting of extreme differences in mothering and fathering suggests that most children do not answer mothering and fathering items in a highly discrepant manner. Thus, high scores on a discrepancy scale may confidently suggest a child's exaggeration of differences in parenting competence by parent gender. Final item inclusion on this scale was determined through rational identification to include equal representation of items loading on

to each of the four mothering and fathering CBQ subscales of Warmth, Monitoring, Communication, and Discipline. This ensures that the discrepancy scale includes items assessing a variety of parenting styles and skills. The initial development of the Parent Discrepancy scale for the CBQ is one of the first of its kind and provides promise that such a scale may be beneficial in determining when children are providing overly discrepant responses between mothering and fathering behaviors.

Data analyses support the development of an Inconsistency scale. Both rational and empirical identification of item-pairs were used to determine 10 item-pairs on the final scale. Similar criterion for empirical identification of ten item pairs has been used on previous measures such as the NEO-PI-R to develop inconsistency scales (Morasco et al., 2007). The development of the Inconsistency scale of the CBQ is unique in that it used both empirical and rational identification of items, rather than either approach alone. It is worth noting that only 3 of the 10 identified item-pairs of the Inconsistency scale consisted of oppositely worded pairs. As the CBQ continues to evolve, future changes in the questionnaire may consider re-wording some of the items to contain more oppositely worded items. This demonstrates another way in which the development of validity scales reveals the nature of the CBQ as a measure.

Development of the Reading Comprehension scale was the only scale in which the addition of new items was previously determined to be necessary. Seven new items were worded to be consistent with existing CBQ items to construct the Reading Comprehension scale. These new items represent over 12% of the CBQ total items. Items with different Flesch-Kincaid ratings between 0.6 and 6.4 were included allowing for a variety of items to represent a respondent's level of reading comprehension on the CBQ. Because each of these

items is a new addition to the scale, data on the nature and content meaning of this scale is not yet available. Future studies are needed to test the usefulness of this scale and to determine cut-off criteria for the Reading Comprehension scales.

Though results supported the development of four out of the five proposed validity scales of the CBQ, data did not support the creation of a Positive Parenting Presentation scale. For the PPP subscales, only two items met empirical criteria for the co-parenting scale, four for the fathering scale and none for the mothering scale. Because of the limited number of items meeting criteria, it was determined that the current items and criteria were not sufficient to create a PPP scale for the CBQ.

The low frequencies of over-reporting positive parenting and co-parenting may be interpreted in several ways to better understand responses on the CBQ. First, the lack of low frequencies in reporting positive parenting and co-parenting may suggest that children generally tend to report their parents in a positive manner. Thus, it may be difficult to separate out children who are over-reporting positive parenting and co-parenting from those who are not. Additionally, these findings may highlight the sensitivity of such items on the CBQ. The nature of the measure may not be sensitive enough to permit a large range of responses and thus children may generally score at the extreme distribution for specific items. As the CBQ evolves, re-wording of questions or response scales may be necessary for the CBQ to evolve into a sensitive enough measure to allow for detecting overly positive reports of parenting and co-parenting. For example, the CBQ may benefit from items that are worded in a neutral manner to pick up on more subtle endorsements of positive parenting. This would potentially allow for the development of PPP scales because extreme scores in

the positive direction may be less likely with neutral wording. Thus, a PPP validity scale may be used to differentiate normative responding versus overly positive reports of parenting.

Overall, Study 1 was successful in identifying threats to valid responding on the CBQ and in the subsequent development of strategies to detect these threats using steps 1 and 2 of the General Validity Scale Model. The use of both empirical and rational identification of items is seen as a major strength of the development of the validity scales. This study was also successful in using study findings to highlight the nature of the CBQ and how this interrelates to the development of each validity scale.

Though this study was a significant first step towards the development of validity scales to identify biased responding, other studies are needed to complete and refine validity scale development under the steps of the GVS Model. Step 3 of the model involves testing validity scales to determine their success in identifying misrepresentations on the CBQ. Completing this step is the focus of Study 2.

Methods – Study 2

Participants

Participants included a convenience sample of 200 undergraduate psychology majors enrolled in a Psychology 101 course at a southeastern university. Participants had the option to participate in the study to fulfill a requirement for completion of the course. Participant's ages ranged from 17 to 50 years old ($M = 20.06$), however participants aged 30 and above were not included in the data analysis resulting in the deletion of three cases and a total of 197 participants. The sample consisted of 95 Caucasians, 44 African Americans, 21 Asians, 12 Multi-racial individuals, 4 Middle Easterners, and 3 Hispanics. Fifty-seven percent of participants reported that their parents are still married, 32.4% reported their parents are

divorced, 4.9% reported that their parents were never married, while 3.8% reported their parents are currently separated.

Measures

Co-parenting Behavior Questionnaire-Revised (CBQ-R). The CBQ-R is a revision of the original CBQ that includes the four newly developed validity scales created in study 1 including the Negative Parenting Presentation, Parenting Discrepancy, Inconsistency, and Reading Comprehension scales. The CBQ-R has retained all 86 items of the original CBQ with the addition of 7 new items that were added for the Reading Comprehension scale. Sample items on the CBQ include “8. I feel caught between my parents”, “46. My mom knows what kinds of things I do after school”, and “81. My dad talks to me about my friends”. *Appendix B* includes each of the CBQ-R items including the 7 additional reading comprehension items.

Student Demographic Form. The Student Demographics Form is a brief 11-item report of standard demographic information including gender, age, and race of the participating student. This form also required students to answer a few questions specifically related to their parents’ marital status and/or divorce including items such as “6. Are your biological parents separated or divorced?” and “11. What were the living arrangements when your parents divorced”. The Student Demographic Form is displayed in *Appendix C*.

Procedures

College students were recruited for the study through the use of Sona Systems, an online survey tool, and individuals decided to participate in the study after reading a brief study description. Students signed up for one of three groups testing different aspects of the CBQ validity scales and completed a corresponding questionnaire packet online which

included the student demographics form and two administrations of the CBQ. Participants read a consent form that explains the procedures of the study and their rights as a subject before agreeing to participate in the study.

Each participant completed the CBQ-R twice with the only difference in the two administrations being the instruction set they were presented with. Prior to completing the CBQ, every student in each group received general instructions that stated the following:

“General Instructions: You are about to complete the same questionnaire twice. The questionnaire asks about you about how you were parented by your mother and father. The only difference between the two questionnaires you will complete are the directions. Please read the directions carefully and then answer the questionnaire once. When you are done, please read the second set of directions and answer the questionnaire a second time.”

After reading the general instructions, every participant completed the CBQ once under standard instructions asking them to complete the CBQ-R truthfully and accurately. These instructions read as follows:

“On the following pages, you will see sentences that have to do with you and your parents. Following each statement, there is a scale from 1 to 5 (1 = almost never, 3 = sometimes, and 5 = almost always). Circle the number that tells how often this statement ACTUALLY happens.”

Completing the CBQ-R more than once involved each participant receiving one of three specific instruction sets asking them to complete the CBQ-R in a specified biased manner. This completion of the CBQ-R involved a group of 69 participants answering the CBQ-R when instructed to attempt over-reporting negative mothering, fathering, and co-

parenting behaviors. This sample was used to validate the NPP scale. A second group of 67 students answered the CBQ-R in a way in which they made an effort to present mothering behaviors in a significantly more positive manner than fathering behaviors. A last set of 64 participants responded to CBQ items in an attempt to show mothering behaviors in a significantly more negative manner than fathering behaviors. These last two groups were used to validate the Parenting Discrepancy scale. The instruction sets given to subjects included one of the following based on the condition they participated in:

1) “INSTRUCTIONS: On the following pages, you will see sentences that have to do with you and your parents. Please answer each of the items about your parents as if you are exaggerating the negative behaviors of your parents. In other words, you want to portray your parents in an unusually bad light. You are trying to answer each item while acting as if both your mother and father show poor parenting skills. Following each statement, there is a scale from 1 to 5 (1 = almost never, 3 = sometimes, and 5 = almost always). Circle the number that tells how often this happens, but remember to act as if you are answering these questions to make your mother’s and father’s parenting look poor.”

2) “INSTRUCTIONS: On the following pages, you will see sentences that have to do with you and your parents. Please answer each of the items about your parents as if you are exaggerating the positive behaviors of your mother while exaggerating the negative behaviors of your father. In other words, you are trying to portray your mother in an unusually positive light and your father in an unusually bad light. You are trying to answer each item while acting as

if your mother shows excellent parenting skills and your father has very poor parenting skills. Following each statement, there is a scale from 1 to 5 (1 = almost never, 3 = sometimes, and 5 = almost always). Circle the number that tells how often this happens, but remember to act as if you are answering these questions to make your mother's parenting look positive and your father's parenting look negative."

3) "INSTRUCTIONS: On the following pages, you will see sentences that have to do with you and your parents. Please answer each of the items about your parents as if you are exaggerating the positive behaviors of your father while exaggerating the negative behaviors of your mother. In other words, you are trying to portray your father in an unusually positive light and your mother in an unusually bad light. You are trying to answer each item while acting as if your father shows excellent parenting skills and your mother has poor parenting skills. Following each statement, there is a scale from 1 to 5 (1 = almost never, 3 = sometimes, and 5 = almost always). Circle the number that tells how often this happens, but remember to act as if you are answering these questions to make your father's parenting look positive and your mother's parenting look negative."

Each participant signed up for one of the three conditions examining behaviors of their parents. Each condition represented one of the three specific instruction sets: 1. Over-reporting negative parenting behaviors, 2. Over-reporting positive mothering while under-reporting positive fathering behaviors (discrepant in mother's favor), or 3. Over-reporting positive fathering while under-reporting positive mothering behaviors (discrepant in father's

favor). Student participants were not aware of which specific biased instruction set they were to receive prior to signing up for the study. Students were also not able to sign up for this study twice. In this way, students were not allowed to participate in more than one of the specific instruction groups. Additionally, the order in which each participant completed the two versions of the CBQ (standard vs. specific instruction set) was randomly counterbalanced in a way in which about half of the participants completed the CBQ under standard instructions first and completed the CBQ under biased instructions second. The other half completed the CBQ under biased instructions first and the standard instructions second. The Sona on-line system for which students completed the study allowed for randomization of surveys in which the two CBQ questionnaires were presented in a counterbalanced fashion.

The procedure used to test each of the validity scale scores at the completion of data collection is outlined below.

Testing the Negative Parenting Presentation scales. Negative Parenting Presentation raw scale scores were calculated for each of students participating in the NPP instruction set. Scores were calculated from the three different subscales of the NPP including mothering, fathering, and co-parenting scales. Additionally, scores were determined for both the standard instruction condition as well as the biased instruction condition for each sample. Thus, every participant had 6 validity scale scores: two NPP-Mothering, two NPP-Fathering, and two NPP-Co-parenting scores. To calculate raw scale scores for each of the three NPP subscales, the sum was computed for item responses of each item determined to be on the specific NPP subscale. For example, every participant's responses (measured on a scale of 1 to 5) were summed for items 1, 3, 4, 5, 8, 10, 14, and 22

to calculate the NPP-Co-parenting scale because these were the items comprising the NPP-Co-parenting scale as determined in Study 1. Before summing responses loading onto the NPP-Fathering and NPP-Mothering subscale, item responses were reverse scored. This was done because the wording of the fathering and mothering items were in the positive direction in which higher numbered responses meant a more positive report of the parents. Reverse scoring these item responses allowed higher numbers on the scale to represent less positive and more negative reports on the NPP-Fathering and NPP-Mothering scales. Thus, when these reverse scores were summed, higher raw scale validity scores on the NPP-Fathering and NPP-Mothering scales represented a more negative presentation of fathering and mothering, respectively, on the CBQ. Raw scale scores from the standard instruction condition were compared to scores from the biased instruction condition for each of the three NPP subscales.

Testing the Parenting Discrepancy scale. Data from two different conditions were used to test the Parenting Discrepancy scale. In particular, one group was given specific instructions to present discrepancies in mother's favor, and another group was instructed to present discrepancies in father's favor as described above. Parenting Discrepancy scale raw scale scores were calculated for the students participating in the discrepant in mother's favor condition and for the students in the discrepant in father's favor condition. Mothering and fathering item pairs identified in Study 1 comprised the Parenting Discrepancy validity scale. Discrepancy raw scale scores were calculated by taking absolute value of the difference in the mothering item response and subtracting them from the corresponding fathering item response. For the "discrepant in mother's favor" conditions, comparisons were made between raw scale discrepancy scores under the standard instruction condition and the raw scale

scores under the biased instruction condition when instructed to answer items in the mother's favor. Comparisons were also made in the same way to test the discrepancy scales under the "favoring father" condition. Thus, discrepancy scores were compared for the students completing the CBQ under standard instructions to scores from the same students completing the CBQ under biased instructions to respond in the father's favor.

Testing the Inconsistency scale. To test the Inconsistency scale, CBQ item responses answered under the standard instruction set of all participating students (students from each condition) were used. CBQ responses of all participants in the standard instruction set were used to calculate a raw scale score on the inconsistency scale. This raw scale score was calculated by taking the absolute value of differences in CBQ numbered responses between each item comprising an item pair on the Inconsistency scale as identified in Study 1. For example, the difference in response scores was computed for the item "14. My parents argue in front of me" and for the item "35. My parents yell at each other" because this comprises an item pair for the inconsistency scale. One item from each item pair on the oppositely worded item pairs on the inconsistency scale was reverse scored. For example, the item "22. When my parents talk to each other, they get angry." was reversed scored while the item "34. My parents get along well." was scored in the original direction. This was done for opposite word pairs so that consistent with same worded item pairs, lower discrepancies in responses would reflect consistent reports on the CBQ while higher numbers in discrepancy scores would represent inconsistent responses on the scale. Once discrepancy scores were calculated for each item pair, those scores were summed to comprise an inconsistency validity scale score. Next, random response sets of the CBQ-R were computer-generated. The number of cases generated was equal to the number of total participants used in the

dataset collected through the Sona systems. Raw scale scores for the generated data were calculated in the same way as for the students' raw scale Inconsistency scores described above. Inconsistency raw scale scores from the undergraduate students were compared to scores from the inconsistency scores calculated from the computer generated data to determine significant differences in raw scale scores for the two conditions.

Testing the Reading Comprehension scale. Similar to the inconsistency scale, data from the CBQ-R standard instruction set of all students were used to test the Reading Comprehension scale. A Reading Comprehension raw scale score was calculated for each student based on their responses on the new 7 Reading Comprehension items. If an item on the Reading Comprehension scale was answered correctly, a raw scale point of one was added to the scale. Incorrectly answered items were scored as a 0. The total Reading Comprehension scale score was calculated by summing the number of items answered correctly for the 7 items comprising the scale. In a process similar to that used in the Inconsistency scale, random responses on the CBQ-R were computer-generated using SPSS. These computer-generated responses were used to calculate additional Reading Comprehension raw scale scores. Reading Comprehension raw scale scores from all the undergraduate students was compared to raw scale validity scores from the computer generated data to determine significant differences in raw scale scores for the two conditions.

Results – Study 2

A data-cleaning plan was implemented prior to calculating raw scale scores as described above or running analyses to test each validity scale. First, cases in which participants were 30 years of age or older were deleted from the dataset. This criterion resulted in the deletion of three cases, reducing the total number of participants to 197. This

was done because these cases were not believed to be accurate appraisals of parenting or co-parenting behavior. This may be because their parents had divorced years earlier and their views on their parents' divorce and their perspective on their parents' overall behaviors had since changed.

Next, a missing data analysis was performed separately for all data in the NPP scale condition and in each of the two Parenting Discrepancy scale conditions. This was done to determine the pattern of missing data within each database to inform the best way to treat missing data. Analyses revealed that missing data within each database was MCAR (Missing Completely at Random) meaning that missing data values were scattered randomly throughout the database and thus posed less of a threat to data analyses (Tabachnick & Fidell, 2007).

Third, frequencies of each participant were examined to determine those cases in which more than 5% of data were missing for a given case. Tabachnick and Fidell (2007) suggests that when about 5% or less of data points are missing for cases than problems with missing data are less serious; however, the opposite is true for cases missing 5% or more of data. Thus, case-wise deletion was completed for those cases in which more than 5% of data were missing. This criterion resulted in the deletion of 8 out of 69 cases being deleted from the NPP database, 4 out of 65 cases being deleted from the Parenting Discrepancy database in which students were instructed to favor their mother, and 6 out of 63 cases being deleted from the Parenting Discrepancy database favoring father. It is important to note that a missing data analysis was performed for only those items comprising the individual validity scales as opposed to all items on the CBQ; however, this resulted in the deletion of no additional cases based on the 5% rule and thus it did not further affect data cleaning.

To address missing data for the remaining cases, missing data was estimated for each dataset. Tabachnick and Fidell (2007) recommend expectation maximization (EM) as a technique for replacing missing data over prior knowledge or mean substitution methods. EM is a more sophisticated data replacement approach that involves creating a missing data correlation matrix and basing inferences about the partially missing data on the likelihood of the distribution of the missing data. EM methods were used in SPSS to impute missing data points. Data sets were then used to test each validity scale created in Study 1. Results of tests performed to validate each validity scale can be found in Table 12.

Table 12.

Results of T-tests to Validate Each Validity Scale

Validity Scale Tested	Mean Standard Instruction Raw Scale Score	Mean Contrived or Randomly Derived Raw Scale Score	T-value	p-value	eta ²
NPP - Co-parenting	16.90	26.70	-6.56	<0.001	0.42
NPP – Mothering	12.69	21.10	-5.66	<0.001	0.34
NPP-Fathering	20.67	23.62	-5.04	<0.001	0.30
Parenting Discrepancy (Favoring Mother)	7.71	13.50	-4.67	<0.001	0.27
Parenting Discrepancy (Favoring Father)	8.46	13.49	-3.52	<0.001	0.18
Inconsistency Scale	13.04	7.06	-16.47	<0.001	0.60
Reading Scale	3.49	1.42	11.91	<0.001	0.44

To test the hypothesis that NPP subscale scores would be significantly lower for the standard instruction set than they would be for the biased instruction set, three correlated

group t-tests were performed. The dependent variable was raw scale scores for each of the three NPP subscales that were calculated as described above. A separate t-test was run to analyze each NPP subscale. For example, a correlated groups t-test was run for the NPP-mothering scale by comparing raw scale validity scores for students when completing the CBQ under the standard instructions with raw scale validity scores for students completing the CBQ under the biased instruction set to over-report negative co-parenting behaviors. Correlated groups t-tests were run in a similar manner to test the validity of the NPP-fathering and NPP-Co-parenting scales.

Results of the correlated groups t-test for the NPP Co-parenting subscale support the hypothesis that validity scales may be used to differentiate normative responding from when an individual reports overly negative co-parenting behaviors, $t(60) = -6.56$, $p < 0.001$, $\eta^2 = 0.42$. Participants scored significantly higher on the NPP validity scale when instructed to over-report negative co-parenting behaviors ($M = 26.70$) than when participants were given standard CBQ instructions ($M = 16.90$). Correlated t-test results for the NPP-mothering subscale also support the hypothesis that this scale can differentiate normative responding from overly negative reports of mothering, $t(60) = -5.66$, $p < 0.001$, $\eta^2 = 0.34$. Participants' scores on the NPP Mothering validity scale were significantly greater when instructed to report mothers in an overly negative manner ($M = 21.10$) than when given standard CBQ instructions ($M = 12.69$). Lastly, results supported the hypothesis that the NPP-fathering subscale can be used to detect exaggerated reports of fathering behaviors from more accurate, normative responding of fathering behaviors, $t(60) = -5.04$, $p < 0.001$, $\eta^2 = 0.30$. Participants' raw scale NPP-fathering scores were significantly higher when instructed to exaggerate negative fathering behaviors ($M = 23.62$) than when given standard instructions to

complete the CBQ ($M = 20.67$). Results for testing the Negative Parenting Presentation scales are displayed in Table 12.

Two correlated groups t-tests were run to validate the Parenting Discrepancy scale. Independent variables were the instruction set condition of each participant and dependent variables were raw scale Parenting Discrepancy scores. The first correlated groups t-test was run to test the hypothesis that the Parenting Discrepancy scale can be used to differentiate normative responding from when an individual reports in a discrepant manner in which mothers are presented more positively than father's. Results supported this hypothesis, $t(60) = -4.67, p < 0.001, \eta^2 = 0.27$ (See Table 12). Parenting Discrepancy scores were significantly higher when participants were given biased instructions to respond in the mother's favor ($M = 13.50$) than when participants were given standard instructions for completing the CBQ ($M = 7.71$). A second correlated groups t-test was run to test the hypothesis that the Parenting Discrepancy scale can also detect discrepant in father's favor responses from normative responding on the CBQ. Results also supported this hypothesis, $t(56) = -3.527, p < 0.001, \eta^2 = 0.18$. Specifically, Parenting Discrepancy scores were higher when participants were provided with biased instructions to favor fathers ($M = 13.49$) than when provided with standard instructions ($M = 8.46$).

A correlated groups t-test was also used to test the hypothesis that randomly generated responses on the CBQ would lead to higher scores on the Inconsistency scale than when participants completed the CBQ under standard instructions. Raw scale validity scores were calculated for both randomly generated responses and for every participant in Study 2. This served as the dependent variable in the analysis. Support for the ability of the Inconsistency scale to detect inconsistent responding was found, $t(178) = -16.47, p < 0.001$,

$eta^2 = 0.60$. Inconsistency scale scores were significantly higher when calculated from random computer generated responses ($M = 13.04$) than when participants were given standard instructions for completing the CBQ ($M = 7.06$). Table 12 displays the results of testing the Inconsistency scale.

A last t-test was performed to test the hypothesis that Reading Comprehension scores would be higher when scores were calculated from randomly generated responses than when calculated from data of participants completed under standard CBQ instructions. Raw scale Reading Comprehension scores were calculated for both conditions as described above; t-test results supported the hypothesis for testing the Reading Comprehension scale, $t(178) = 11.91, p < 0.001, eta^2 = 0.44$. Specifically, scores on this scale were significantly lower for the randomly generated response condition ($M = 1.42$) than for the standard instruction condition ($M = 3.49$). See Table 12 for results.

Discussion – Study 2

Study 2 demonstrated the success and potential significant benefit in using the Co-parenting Behavior Questionnaire validity scales to detect inaccurate responding on the measure. This was accomplished by employing step 3 of the General Validity Scale (GVS) Model to test the separate validity scales created in Study 1. The Negative Parenting Presentation scales for Co-parenting, Mothering, and Fathering, as well as the Parenting Discrepancy Scale, were tested using a within group simulation design. This design has been used with success in prior studies testing validity scales (Baer & Miller, 2002). Raw scale validity scores for participants completing the CBQ under standard instruction were compared to the participants' raw scale validity score when asked to complete the CBQ in a specified biased manner. The Inconsistency scale and Reading Comprehension scale were

both tested by comparing validity scale scores for participants completing the CBQ under standard instructions to validity scale scores computed from randomly generated computer responses.

Results for testing the NPP validity scales support the hypothesis that these scales may be used to accurately detect distorted responses reflecting overly negative parenting or co-parenting on the CBQ. Specifically, the NPP Co-parenting scale scores were significantly higher when participants were instructed to answer the CBQ as if they were presenting both of their parents in a negative manner than when participants completed the CBQ under standard instructions. The same was found true for the NPP Mothering and NPP Fathering scales. Thus, raw scale scores were higher when instructed to present mothers (NPP Mothering) or fathers (NPP Fathering) in a negative manner than when participants completed the CBQ using standard instructions. These findings support the contention that NPP validity scales may be used to discriminate between normative responses and biased reporting on the CBQ. For example, if a child who is completing the CBQ scores very high on the validity scale score for NPP-Mothering, the child is likely exaggerating negative mothering behaviors to an extent significantly greater than the normative population. The validity scale could detect this biased responding, whether intentional or not, and serve as an alert for the test examiner.

One interesting finding from testing the NPP scales was the seemingly large raw scale score for the NPP-Fathering scale for the standard instruction set. The mean for this raw scale score was approximately 4 points higher than the average NPP Co-parenting score and almost 8 points higher than the average NPP-Mothering score. Additionally, although significant differences were found between the NPP-Fathering scores under standard

instructions versus biased instructions, there was only about a 3 point difference on average between these scale scores. These findings suggest that on average individuals are more likely to present their fathers in a more negative manner on the CBQ-R than their mothers. There may be several reasons for this finding including that fathers may generally be less involved with their children than mothers in the American family structure. Fathers may also be rated more negatively post-divorce due to custody arrangements that often result in children spending less time with fathers than mothers. Additionally, there may be greater overall variability in fathering behaviors than mothering behaviors resulting in proportionately greater positive ratings. Lastly, this finding may also speak to the nature of the CBQ and the potential lack of sensitivity of the measure in rating fathering behaviors. Future research to create scoring criteria and determine validity scale norms may prove helpful in further understanding potential reasons for such differences in NPP validity scale scores.

Support was also provided for the ability of the Parenting Discrepancy scale to detect large reported discrepancies between mothering and fathering behaviors. Two tests were performed to validate the discrepancy scales. One test examined scale scores in which mothers were presented in a positive manner while fathers were presented in a negative light and the other test used data in which fathers were presented in a positive light while mothers were presented negatively. Parenting Discrepancy scores were significantly higher for contrived scores (favoring mother and favoring father) than scores produced when participants completed the CBQ when given standard instructions. These findings suggest that the Parenting Discrepancy scale can successfully detect when a child is reporting one parent much more positively (or negatively) than the other parent. Additionally, the scale is

able to detect these discrepancies regardless of whether a child is presenting their mother or father in an unusually positive manner when compared to the other parent.

Support for the hypothesis that the Inconsistency scale can be used to detect inconsistent or random responding on the CBQ was also found. Specifically, Inconsistency raw scale scores were significantly higher when computed from a randomly generated dataset than when calculated from a normative dataset under standard instructions. These findings indicate that the Inconsistency scale may successfully reveal random responding on the CBQ.

By nature, the CBQ-R Inconsistency scale may detect both intentional random responding as well as unintentional inconsistent responding resulting from low reading level; however the Inconsistency validity scale alone cannot detect the reason for which this random responding occurs. By nature, the Reading Comprehension scale may help to inform the test examiner whether random responding is in fact due to low reading level. Tests to validate and provide support for such use of the Reading Comprehension scale were also performed in Study 2.

The hypothesis that a Reading Comprehension scale could be created to assess a child's ability to understand current items on the CBQ was supported. Participants completing the CBQ under standard instructions got more items correct and, thus, had significantly higher Reading Comprehension scores than validity scale scores computed from a randomly generated data set. This suggests that the Reading Comprehension scale may indeed be used to better understand if potential random responding is due to poor reading comprehension. When examining sample means, the reading scale score (out of a possible 7 points) was only about 2 points higher for the standard instruction set than for the contrived

set. This was an unexpectedly low reading scale score mean when considering the education level of the sample.

Item level frequencies for each reading scale item were examined to determine if more individuals received a correct score for items that were rated as easier on the Flesch-Kincaid scale than for more difficult items. No such pattern was found in the data, thus, the level of item difficulty did not affect the correctness of the item. These results may suggest that low Reading Scale scores were not due to education level, but perhaps were due to the nature of the population sampled. This convenience sample of college students may have been less involved in answering the CBQ and was instead more interested in completing the measure to obtain course credit. Other hypotheses about this sample include that they may have simply answered these Reading Comprehension scale items in a more random manner than other CBQ items because these items required more thought resulting in a correctness score rather than an opinion or the respondent's perspective. Additionally, because these items are located at the end of the CBQ, raters may have guessed items incorrectly due to fatigue from completing the questionnaire. CBQ Reading Scale score items may need to be placed at the beginning of the measure or embedded throughout the measure to correct for this issue.

General Discussion

The current studies significantly contributed to the needed movement towards the use of validity scales in assessment tools used with children in which important clinical and forensic decisions are made. The creation of the General Validity Scale (GVS) Model and development of validity scales for the Co-parenting Behavior Questionnaires are two distinct contributions of the current study. The GVS Model was found to be empirically supported

and useful when applying the model to develop CBQ validity scales. Additionally, the studies support the development and use of several CBQ validity scales to successfully detect invalid responding. Study 1 employed steps 1 and 2 of the GVS Model by 1) identifying possible validity threats of the CBQ and 2) developing strategies to detect the threats in question. Study 2 accomplished step 3 of the GVS Model by testing the CBQ validity scales developed in Study 1.

Five possible threats to validity were initially identified in Study 1, including children presenting parenting and co-parenting in an unusually negative manner, reporting overly positive parenting and co-parenting behaviors, reporting in a highly discrepant manner between mothering and fathering behaviors, responding inconsistently or randomly on the CBQ, and inaccurately reporting due to a low reading level. Both empirical and rational identification of items was used to determine which items loaded onto the different possible CBQ validity scales. Support was found for the development of four different validity scales including the three different subscales of the Negative Parenting Presentation scales (NPP Co-parenting, mothering, and fathering), the Parenting Discrepancy scale, the Inconsistency scale, and the Reading Comprehension Scale. Findings did not support the development of Positive Parenting Presentation subscales due to the limited number of items meeting predetermined criteria for the scales.

Study 2 was successful in testing each of the four validity scales. Validation of these scales involved determining if raw validity scale scores were significantly different for scores calculated from a standard administration of the CBQ versus scores calculated from contrived data. Raw scale scores for the NPP scales and Parenting Discrepancy scores were significantly higher when participants were asked to intentionally bias responses in a

specified manner than when participants completed the CBQ when provided with standard instructions. The inconsistency scales were also deemed valid after showing that validity scores were significantly higher when scores were calculated from randomly generated data than when determined from CBQs administered under standard directions. Lastly, there were a significantly greater number of correct items on the Reading Comprehension scale for standard instruction administrations of the CBQ-R than there were on a randomly generated dataset.

Implications in Clinical and Forensic Settings

Validity Scales of the Co-parenting Behavior Questionnaire-Revised show great promise in clinical and forensic settings in which important decisions are made that affect children and families from divorced backgrounds. Specifically these scales may be used to determine when and how children are responding in a biased, inaccurate manner on the CBQ-R. Professionals using the CBQ-R may use validity scales to determine whether to identify the CBQ-R as invalid and thus disregard testing scores, retain the scores but interpret them with caution, or use the CBQ-R to identify psychopathological influences that may have lead to invalid data.

In clinical settings, treatment providers working with children and families from divorced backgrounds may opt to administer the CBQ-R as larger test battery to aid in treatment planning. When CBQ-R profiles are determined to be valid, a clinician may confidently use CBQ-R results as part of the clinical process. If specific validity scales for the CBQ are deemed invalid, a clinician may decide to disregard or “throw out” information obtained from this measure. The clinician may also choose to use information presented in the CBQ-R results but to interpret these findings with caution. A last way clinicians may

choose to use the CBQ-R when presented with invalid results is to identify which scales are invalid to make hypotheses about why a child is responding in a biased manner. For example, if a child scores extremely high on the Parenting Discrepancy validity scale this may indicate that one parent is alienating a child from his or her ex-spouse thus CBQ-R ratings are presented as more negative for the alienated parent. These high scores could also imply that one parent is seen as a fun or permissive parent while the other parent is more of a disciplinarian or strict parent. In either case, clinicians may use these test results from the CBQ-R to make hypotheses about why a child is responding in a biased manner and may further inform the assessment process.

CBQ-R validity scales are necessary and urgent in forensic settings in which major decisions regarding divorce or child custody is in question. Courts are required to weigh the importance of any presented data to make decisions in forensic settings. Weighing the accuracy or the validity of data is an important step when presenting data to the Court that is both credible and objective. Potential uses of the CBQ-R in forensic settings include professionals using the measure as part of the determination of child custody. In such cases decisions of the professional and judge will greatly benefit from knowing if a child is providing inaccurate responses to the CBQ-R. If results of the CBQ-R are deemed valid, those involved in the decision making process may feel more confident in using the measure as part of their larger assessment battery during custody evaluations. However, if deemed invalid, professionals may decide to disregard CBQ-R results or interpret them with caution since they may not paint an accurate picture of parenting and co-parenting behaviors. Lastly, invalid scale scores on the CBQ-R may be used to better understand possible reasons for why a child is responding in a biased manner. For example if a child scores in the invalid range on

the Inconsistency scale of the CBQ, the test examiner may hypothesize that this is occurring because the child does not want to feel caught between his or her parents and wishes to remain out of the divorce process. This information may inform the professional as to what further assessment is needed to better understand important processes occurring within the divorced family.

The current study moves beyond the development of CBQ-R validity scales to also develop and test the use of the General Validity Scale Model as an important guide in validity scale development. No standard model is currently available as an aid in creating validity scales; thus the GVS Model is the first of its kind. The General Validity Scale Model presents three major steps to guide validity scale development for any given measure. These steps include 1) identify possible threats to validity for the measure in question, 2) develop strategies to detect these threats, and 3) test and norm the created validity scales. Potentially any measure may follow these steps to create validity scales specific to the assessment process.

Measures of child report in particular may benefit from following the GVS Model to create scales to detect invalid responding. The assessment field often makes assumptions that children can accurately complete measures of self-report or reports on other's behaviors. However there are several reasons why children may intentionally or inadvertently respond inaccurately on such measures. These reasons include:

1. Cognitive Capacity. Children simply do not have the cognitive capacity or have not reached the appropriate developmental level required to complete the assessment task. This may result in an inaccurate response style on child-report measures.

2. Disinterest in assessment. Inconsistent responding by children because they are disinterested in the assessment process. Because of their lack of involvement in completing a measure they may simply choose item responses at random.

3. Social Desirability. Some children may misrepresent reports of behaviors to the test examiner, their parents, or others in order to present themselves or others in a socially desirable way. For example, if a child is completing a measure on parenting behaviors as part of a custody evaluation, they may respond in a way they assume their parents(s) would want them to respond.

4. Nature of parent-child relationships. In some cases, children's love and care of their parent's may override their ability to accurately respond to items such as those that assess parenting behaviors.

5. Cry for help. Children may over-report negative behaviors of themselves or their parents as a cry for help. This may be a child's way of communicating to the examiner that problems exist within the individual(s) or family.

The need for validity scales becomes more evident after reviewing the multitude of reasons why invalidity may result. Presently, there are very few child assessment tools with embedded validity scales (Borum & Stock, 1993) despite the need for such scales. By using the GVS Model as a guide, validity scales may be developed, tested, and embedded within a measure to detect when inaccurate responding is occurring.

The use of the GVS Model to guide development of the CBQ-R validity scales marks the first application of the model to the development of validity scales; however, the model may potentially be applied to the development of validity scales for any given measure. The GVS Model was formed after examining numerous measures with already existing validity

scales to determine common ways of creating such scales. The model can be applied in the development of validity scales regardless of whether similar validity scales of its kind already exist. This was demonstrated in the development of the Parenting Discrepancy scale and Reading Comprehension scale of the CBQ-R for which validity scales for other measures did not previously exist. Though the GVS Model holds promise in guiding the development of validity scales, it is an evolving model that may benefit from further development. The model may be elaborated upon to include additional potential common threats to validity, strategies for detecting threats, and methods of testing validity scales that are either frequently used or currently non-existing. Further development of the GVS Model may suggest more specific, sound ways of developing validity scales such as using both empirical and rational strategies for identifying items comprising validity scales.

Limitations and Future Directions

Though the current study shows promise in the use of validity scales for the CBQ-R, it is not without limitations. One such limitation in this study includes that the data used to develop the CBQ validity scales in Study 1 was from a previously collected dataset using responses to the original CBQ containing 92 items. Since that data collection, the CBQ has evolved to an 86 item report of parenting and co-parenting; thus, the current study involved the deletion of 6 CBQ items and responses from the data set to determine item inclusion on each validity scale. Additionally, 7 new items were added to the CBQ-R to include items created for the Reading Comprehension scale. These changes in instrumentation may result in different psychometric properties for the new CBQ-R than the original 92 item version from which validity scales were created. Because the data were not used as part of a longitudinal design involving retesting subjects, this limitation in instrumentation is not

believed to greatly compromise the internal validity of the scales developed in Study 1 (Kazdin, 2003).

The nature of the participants selected for this study poses several threats to validity. One such threat is that convenience samples were used in Study 1 and 2 to create and test the validity scales. Kazdin (2003) discusses samples of convenience as a threat to external validity of study findings because it challenges the generality of study findings. Generalization of study findings from both younger siblings of college students used in Study 1, as well as college students in Study 2, may be threatened due to the specific demographics and characteristics of these participants. An additional limitation related to the nature of participants is that Study 2 data was gathered from participants that were not within the age range for which the CBQ was designed to be used. The original CBQ was intended for use with 10 to 18 year olds from divorced backgrounds. Data used in Study 2 was comprised of participants aged 17 to 30 limiting the external validity of the study findings. Lastly, the sample used included individuals from both divorced and intact families. This serves as a limitation given that the original CBQ was designed and tested on children from divorced families only. Future research should address these limitations by testing the validity scales on a non-convenience sample of 10 to 18 year olds from divorce backgrounds. This will bring clarity to the generalizability of the research and will determine if CBQ-R validity scales can accurately detect threats to validity in this population.

Limitations also include that non-normative samples were not used to develop or test CBQ-R validity scales. As previously discussed, the CBQ-R may be a valuable tool in clinical and forensic settings. The current study limits the generalizability of the ability of the CBQ-R to detect threats of validity in clinical or forensic settings. Thus, future studies are

needed to test CBQ-R validity scales on non-normative samples such as clinical samples of children from divorced backgrounds. Research with clinical and forensic samples will bring understanding about the use of the CBQ-R and further implications of the validity scales for this measure. Several recommendations have been made to address the limitations of the current study. Future research is also needed to further develop the use of validity scales beyond specific limitations of the study.

Recommendations for future development of the CBQ-R include determining norms and cut-off scores for the CBQ-R validity scales. These recommendations are imbedded within step 3 of the GVS Model in which validity scales should be both tested and normed with a population for which the measure was intended. Establishing norms of validity scales will ultimately help to develop scoring criteria for the CBQ-R validity scales and interpretation strategies for the entire measure. Once the CBQ-R is normed with an appropriate population, scores will provide necessary structure for scoring criteria. Developing norms and scoring criteria will ultimately aid in determining appropriate cut-off scores for each validity scale.

Future research is also needed to determine rules for excluding CBQ protocols and to aid in decision making once protocols are deemed invalid. The current study did not develop rules for determining at what point CBQ-R profiles should be considered invalid. Studies to develop norms and scoring criteria will also provide guidance for developing rules for determining invalidity of CBQ-R profiles. This step should follow with determining possible decisions on how to use the CBQ-R once it is deemed invalid. This decision may depend on the context for which the measure is being used. One may decide to “throw out” all CBQ-R test results, interpret the CBQ-R with caution, or use the invalid CBQ-R protocol to better

understand child and family psychopathology. These decisions may look differently for professionals using the CBQ-R in a clinical context than in a forensic context. Further research could help to help guide this decision making process for varied contexts.

The current study was an important contribution and first step in the development of CBQ validity scales. These validity scales may prove especially useful in clinical and forensic context in which decisions on treatment and custody arrangements for children from divorced families may occur. The study also introduced the General Validity Scale Model as a guide in validity scale development for any given measure. The model was applied and proven useful in development for the CBQ validity scales. Although the current study is not without limitations it moves the field forward in demonstrating the use of the GVS Model to create much needed validity scales for child-report measures. It also provides growth and confidence in the assessment of parenting and co-parenting behaviors as is done in clinical and forensic settings in families of divorce.

List of References

- Archer, R. P., Fontaine, J., & McCrae, R. R. (1998). Effects of two MMPI-2 validity scales on basic scale relations to external criteria. *Journal of Personality Assessment, 70*, 87-102. doi:10.1207/s15327752jpa7001_6
- Baity, M. R., Siefert, C. J., Chambers, A., & Blais, M. A. (2007). Deceptiveness on the PAI: A study of naïve faking with psychiatric inpatients. *Journal of Personality Assessment, 88*, 16-24. doi:10.1207/s15327752jpa8801_03
- Baer, R. A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: A meta-analytic review. *Psychological Assessment, 14*, 16-26. doi:10.1037/1040-3590.14.1.16
- Baer, R. A., & Wetter, M. W. (1997). Effects of information about validity scales on underreporting of symptoms on the personality assessment inventory. *Journal of Personality Assessment, 63*, 402-413. doi:10.1207/s15327752jpa6802_10
- Borum, R. & Stock, H. V. (1993). Detection of deception in law enforcement applicants. *Law and Human Behavior, 17*, 157-166.
- Briere, J. (1996). The trauma symptom checklist for children. Florida: Psychological Assessment Resources, Inc.
- Butcher, J.N. (1979). Use of the MMPI in personnel selection. In J.N. Butcher (Ed.), *New developments in the use of the MMPI* (pp. 165-201). Minneapolis: University of Minnesota Press.
- Edens, J. F., & Ruiz, M. A. (2006). On the validity of validity scales: The importance of defensive responding in the prediction of institutional misconduct. *Psychological Assessment, 18*, 220-224. doi:10.1037/1040-3590.18.2.220
- Fricker, A. E., & Smith, D. W. (2001). Trauma specific versus generic measurement of distress and the validity of self-reported symptoms in sexually abused children. *Journal of Child Sexual Abuse, 10*, 51-66. doi:10.1300/J070v10n04_04
- Furr, R. M., & Bacharach, V. R. (2007). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage Publications.
- Greene, R. L. (2000). *The MMPI-2: An interpretative manual (2nd ed.)*. Boston: Allyn and Bacon.
- Kucharski, L. T., Toomey, J. P., Fila, K., & Duncan, S. (2007). Detection of malingering of psychiatric disorder with the personality assessment inventory: An investigation of criminal defendants. *Journal of Personality Assessment, 88*, 25-32. doi:10.1207/s15327752jpa8801_04

- Macie, K. M., & Stolberg, A. L. (2003). Assessing parenting after divorce: The Co-parenting Behavior Questionnaire. *Journal of Divorce & Remarriage*, 39, 89-107. doi:10.1300/J087v39n01_06
- Medoff, D. (1999). MMPI-2 Validity scales in child custody evaluations: Clinical versus statistical significance. *Behavioral Sciences and the Law*, 17, 409-411. doi:10.1002/(SICI)1099-0798(199910/12)17:4<409::AID-BSL357>3.0.CO;2-N
- Morasco, B. J., Gfeller, J. D., & Elder, K. A. (2007). The utility of the NEO-PI-R validity scales to detect response distortion: A comparison with the MMPI-2. *Journal of Personality Assessment*, 88, 276-283.
- Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Mullett, E. K., & Stolberg, A. L. (1999). The development of the Co-parenting Behaviors Questionnaire: An instrument for children of divorce. *Journal of Divorce & Remarriage*, 31, 115-137. doi:10.1300/J087v31n03_07
- Piedmont, R. L. (1998). *The revised NEO Personality Inventory: Clinical and research applications*. NY: Plenum Press.
- Schinka, J. A., Kinder, B. N., & Kremer, T. (1997). Research validity scales for the NEO-PI-R: Development and initial validation. *Journal of Personality Assessment*, 68, 127-138. doi:10.1207/s15327752jpa6801_10
- Schum, L. (2003). Standardization of the co-parenting behavior questionnaire; advancing scoring interpretability through construct validation and normative comparison (Unpublished master's thesis). Virginia Commonwealth University, Richmond, VA.
- Schum, L., & Stolberg, A. L. (2007). Standardization of the Co-parenting Behavior Questionnaire. *Journal of Divorce & Remarriage*, 47, 103-132. doi:10.1300/J087v47n03_06
- Stolberg, A. L., Ferrante, J., & Schum, L. (2006) Conceptual and Empirical Support for the Measurement of Co-Parenting. In *Forum on Public Policy: Child Psychology*. Oxford, England.: Oxford University Press.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using Multivariate Statistics (5th ed.)*. Boston: Allyn and Bacon.

Appendix A

Item Loadings of the CBQ by Subscale

Item numbers of the CBQ loading on to co-parenting subscales:

Triangulation: 2, 4, 7, 8, 9, 10, 11, 15, 16, 18, 25, 27

Parental Conflict: 1, 3, 5, 6, 14, 21, 22, 31, 34, 35

Parental Communication: 17, 19, 20, 23, 24, 29

Parental Respect/Cooperation: 12, 13, 26, 28, 30, 32, 33, 36

Item numbers of the CBQ loading on to parenting subscales:

Mother Warmth: 41, 44, 47, 48, 65, 81*, 82

Mother Discipline: 40, 53, 57, 64, 72, 81*, 84, 86

Mother-Child Communication: 38, 39, 59, 63, 78, 80

Mother Monitoring: 45, 46, 58, 60, 75

Father Warmth: 37, 43, 50, 62*, 66, 69, 71

Father Discipline: 49, 61, 62*, 68, 74, 76, 79, 83

Father-Child Communication: 42, 54, 70, 73, 77, 85

Father Monitoring: 51, 52, 55, 56, 67

*Loads on multiple scales

Appendix B

Co-parenting Behavior Questionnaire - Revised

1. My parents complain about each other.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

2. My dad tells me bad things about my mom.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

3. My parents argue about money in front of me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

4. When my parents argue, I feel forced to choose sides.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

5. When my parents talk to each other, they accuse each other of bad things.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

6. My parents talk nicely to each other.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

7. My mom asks me questions about my dad that I wish she would not ask.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

8. I feel caught between my parents.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

9. My dad asks me to carry messages to my mom.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

10. My parents fight about where I should live.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

11. My dad asks me questions about my mom that I wish he would not ask.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

12. My mom wants me to be close to my dad.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

13. When my mom needs to make a change in my schedule, my dad helps.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

14. My parents argue in front of me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

15. My mom tells me to ask my dad about child support.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

16. It is okay to talk about my mom in front of my dad.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

17. My parents talk to each other about my problems.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

18. It is okay to talk about my dad in front of my mom.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

19. My parents talk to each other about how I feel about the divorce.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

20. My parents talk to each other about my school and my health.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

21. My dad gets angry at my mom.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

22. When my parents talk to each other, they get angry.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

23. My parents talk to each other about big choices in my life.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

24. My parents talk to each other at least once a week.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

25. My mom tells me bad things about my dad.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

26. When my mom needs help with me, she asks my dad.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

27. My mom asks me to carry messages to my dad.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

28. My mom tells me good things about my dad.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

29. My parents talk to each other about the good things I do.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

30. When my dad needs help with me, he asks my mom.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

31. My mom gets angry at my dad.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

32. My dad tells me good things about my mom.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

33. My dad wants me to be close to my mom.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

34. My parents get along well.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

35. My parents yell at each other.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

36. When my dad needs to make a change in my schedule, my mom helps.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

CBQ - PART B

37. My dad likes being with me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

38. My mom and I have friendly talks.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

39. My mom asks me about my day in school.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

40. When I do something wrong, my mom talks to me about it.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

41. I feel that my mom cares about me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

42. My dad talks to me about big choices in my life.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

43. I feel that my dad cares about me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

44. I spend time doing fun things with my mom.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

45. My mom knows who my friends are and what they are like.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

46. My mom knows what kinds of things I do after school.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

47. My mom likes being with me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

48. I talk to my mom.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

49. I have chores to do at my dad's house.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

50. My dad says he loves me and gives me hugs.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

51. When I leave the house, my dad knows where I am and who I am with.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

52. If I have problems in school, my dad knows about it.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

53. When I break one of my mom's rules, she punishes me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

54. My dad asks me about my day in school.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

55. My dad knows who my friends are and what they are like.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

56. My dad knows what kinds of things I do after school.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

57. I have chores to do at my mom's house.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

58. When I leave the house, my mom knows where I am and who I am with.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

59. My mom talks to me about big choices in my life.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

60. If I have problems in school, my mom knows about it.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

61. When I do something wrong, my dad talks to me about it.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

62. My dad praises me when I do something good at home or at school.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

63. I talk to my mom about my problems.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

64. If I get in trouble at school, my mom punishes me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

65. My mom says nice things about me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

66. I spend time doing fun things with my dad.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

67. My dad knows who my teachers are and how well I am doing in school.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

68. I have rules to follow at my dad's house.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

69. I talk to my dad.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

70. I talk to my dad about my problems.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

71. My dad says nice things about me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

72. I have rules to follow at my mom's house.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

73. My dad and I have friendly talks.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

74. When my dad says he is going to punish me, he does it.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

75. My mom knows who my teachers are and how well I am doing in school.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

76. 76. When I break one of my dad's rules, he punishes me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

77. My dad talks to me about my friends.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

78. My mom talks to me about my friends.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

79. My dad is patient with me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

80. I talk to my mom about things that I do well.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

81. My mom praises me when I do something good at home or at school.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

82. My mom says she loves me and gives me hugs.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

83. If I get in trouble at school, my father punishes me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

84. My mom is patient with me.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

85. I talk to my dad about things I do well.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

86. When my mom says she is going to punish me, she does it.

1	2	3	4	5
Almost Never		Sometimes		Almost Always

87. “*My mom and I have friendly talks.*” means:

- A) My mom and I talk about things that make me feel really good.
- B) My mom and I talk about problems I’m having with friends.
- C) My mom and I enjoy cooking together.
- D) My mom and I argue a lot.
- E) My mom and I talk and then I feel bad.

88. “*When I leave the house, my dad knows where I am and who I am with.*” means that:

- A) When I am not at home my dad knows who I am hanging out with.
- B) When I leave my house, my dad knows what time I will be home.
- C) When we go out, my dad spends time doing fun things with me.
- D) When I am at a friend’s house, my dad does not know where I am or what we are doing.
- E) When I am out, my dad has trouble keeping up with where I am.

89. “*When I break one of my dad’s rules, he punishes me.*” means that:

- A) When I do something that my dad does not want me to do, I get in trouble for it.
- B) When I do not do something that is important to my dad, he gets upset.
- C) When I break a rule at my dad’s house, he thinks my mom will punish me.
- D) When I do something my dad likes, he usually lets me know.
- E) When I do something my dad does not want, I can get away with it.

90. *"I feel caught between parents."* means that:

A) I feel like I get stuck in the middle of my parents' fights and feel like I have to choose sides.

B) I feel like my parents like to fight with each other a lot.

C) I feel like my mom and dad both do fun things with me.

D) I feel like both of my parents want to know my opinion on things even if I do not agree with what they want.

E) I feel like both of my parents really want me to be close to and to spend a lot of time with the other parent.

91. *"When I do something wrong, my mom talks to me about it."* means that:

A) When I don't do what is expected or I get in trouble, my mom and I have a serious talk about it.

B) When I don't try hard at school, my mom gets upset.

C) When I do something wrong at my dad's house, my mom does not care.

D) When I do something good, my mom tells me.

E) When I do something wrong my mom usually does not know about it.

92. *"My parents talk to each other about big choices in my life."* means that:

A) My parents talk to each other about important decisions they make about me.

B) My parents talk to each other when I'm going to a friend's house.

C) My parents yell at each other when they talk about work.

D) My parents blame each other when I have important things going on.

E) My parents do not talk to each other very often.

93. *"My parents complain about each other."* means that:

A) My parents say bad things about each other and blame each other for things that happen.

B) My parents really do not like each other.

C) My parents don't take me to school on time.

D) My parents say good things about each other.

E) My parents are patient with each other and listen to each other before making decisions.

Appendix C

Student Demographic Questionnaire

1. What is your gender?
 - a) Male
 - b) Female

2. What is your race?
 - a) White
 - b) Black
 - c) Hispanic
 - d) Asian
 - e) Middle Eastern
 - f) Multiracial

3. What is your religious affiliation, if any?
 - a) Protestant Christian
 - b) Roman Catholic
 - c) Evangelical Christian
 - d) Jewish
 - e) Muslim
 - f) Hindu
 - g) Buddhist
 - h) Other _____
 - i) No affiliation

4. What year/class are you in currently?
 - a) Freshman
 - b) Sophomore
 - c) Junior
 - d) Senior
 - e) Graduate/Professional

5. How old are you? ____ years ____ months; Birth date: __/__/____

6. Are your biological parents separated or divorced?
 - a) No, they are currently married
 - b) No, my parents never married
 - c) Yes they are separated but not divorced
 - d) Yes they are divorced

If answered C or D to #6 please proceed to number 7. If answered no, please move on to the next questionnaire.

7. What was your parents' marital status from the time of your birth up to age 18?

- a) Married
- b) Never married
- c) Separated but not divorced
- d) Divorced

8. If your parents separated at any time from birth through age 18, how long did they separate for? If they were separated multiple times during this time period, please list the lengths of each separation.

_____ years _____ months
_____ years _____ months
_____ years _____ months
_____ years _____ months

9. How old were you when this separation occurred? (If multiple separations occurred please list the ages of each.)

_____ years _____ months
_____ years _____ months
_____ years _____ months
_____ years _____ months

10. i. If your parents separated from the time you were born until you were 18, did the separation(s) lead to a divorce?

- a) No
- b) yes

ii. IF your parents divorced, please list your age when the divorce occurred.

_____ years _____ months

11. What were the living arrangements when your parents were divorced?

- a) All nights with mom
- b) All nights with dad
- c) Most nights with mom, some with dad
- d) Most nights with dad, some with mom
- e) About half the nights with mom, half the nights with dad

Vita

Kimberly Marie Parker was born on September 23, 1984 in Greenville, North Carolina, and is an American citizen. She graduated from D.H. Conley High School, Greenville, North Carolina in 2002. She received her Bachelor of Arts in Psychology and Exercise and Sport Science from The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina in 2006. She subsequently was employed as a research assistant at innovation, Research, and Training, Inc. in Durham, North Carolina between May of 2006 and July of 2008.