Multistudy Report

Development and Validation of the Parental Food Choice Guilt Scale

Haley E. Yaremych[1,2] and Susan Persky[2]

[1]Department of Psychology & Human Development, Vanderbilt University, Nashville, TN, USA
[2]Social & Behavioral Research Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA

Abstract: Guilt is an important correlate of the food choices parents make for their children. However, without a validated self-report measure of feeding-related guilt, researchers remain unable to reliably assess its consequences for parents and children. This study developed and validated the Parental Food Choice Guilt (PFCG) Scale. Items were iteratively developed based on feedback from parents and content experts and refined through three preliminary rounds of data collection. The final 7-item PFCG was evaluated using an Internet-based sample of 294 parents (194 mothers, 99 fathers) of children aged 3–13. Participants responded to the PFCG alongside measures of child eating habits, parent feeding behavior, and guilt-proneness. Exploratory factor analysis indicated the unidimensionality of the underlying construct. A Graded Response Model indicated all items functioned well psychometrically. Internal consistency was high (Cronbach's α = .90). Differential item functioning was not detected by parent gender, parent or child weight status, or parent education. The PFCG was positively correlated with global guilt, parenting-related guilt, and poorer child eating habits, and negatively correlated with feeding self-efficacy and healthfulness of the child's diet. The PFCG is a reliable and valid measure that will aid researchers across many applications, including clarifying guilt's role in child feeding.

Keywords: child feeding, guilt, parents, measurement, item response theory

In behavioral and clinical science, guilt is regarded as one of the most fundamental self-conscious emotions. Extensive study has investigated guilt as a motivator of adaptive behavior and prosocial tendencies in certain contexts, but also as an inhibitor of positive behavior change in other contexts, rendering it among the most complex affective states in psychology (Tangney, 1996). Accordingly, in the realm of health behavior, guilt has been identified as a potential driver of positive dietary changes (Conradt et al., 2008), yet has also been linked to increased unhealthy behavior (Hofmann & Fisher, 2012). In studies of guilt regarding parents' food choices for their children, mixed findings persist, likely due to inconsistent and unvalidated measurement tools used in past work. The goal of the current study is to develop and validate a novel measure of parents' guilt about the food choices they make for their children.

The food choices that parents make for their children are linked to many important outcomes, including children's self-regulation of food intake (Birch et al., 2001), food preferences and attitudes toward eating (Savage et al., 2007), and weight (Birch & Fisher, 2000). Efforts to elucidate the affective and contextual precursors to parents' food decisions have identified guilt as an important construct to consider for mothers (Johnson et al., 2011; Noble et al., 2005; Pocock et al., 2010) and, according to emergent evidence, fathers (Harris et al., 2020). However, it remains unclear whether guilt is helpful or harmful in this domain; guilt may motivate intentions to make better food choices in the future, but may also trigger the self-gratifying behavior of appeasing one's child with less healthy foods (Hagerman et al., 2019). One important barrier to elucidating guilt's role in parents' food choices is the absence of a validated self-report instrument, which motivates the current study.

Theorists have defined guilt as a self-conscious emotion arising from real or imagined wrongdoing or failure, which possesses unique affective and behavioral manifestations. In terms of behavioral tendencies, guilt is thought to be characterized by a desire to "approach and amend," such as by confessing, apologizing, or making reparation (Tangney, 1996). Additionally, guilt is associated with a specific wrongdoing or behavior as the object of negative evaluation, rather than the entire self; in other words, guilt is characterized by "I did a bad thing" rather than "I am a bad person" (Tangney, 1996; Tracy & Robins, 2006).

In general, the self-report instrument with which guilt is assessed is often tied to whether guilt is linked to positive or negative outcomes. Guilt is typically measured in one of two ways. First, *adjective-based scales* contain a checklist of emotions, and participants indicate the extent to which they have experienced each emotion in, for example, the past few weeks (e.g., the Personal Feelings Questionnaire [PFQ]; Harder & Lewis, 1987). Adjective-based scales are thought to capture chronic, maladaptive guilt, as guilt assessed via these scales is predictive of poor outcomes such as low self-esteem (Kim et al., 2011; Tignor & Colvin, 2019). Adjective-based scales do not capture the eliciting situations from which guilt arises; this is problematic given that guilt is theorized to occur due to a specific action. Second, *scenario-based scales* (e.g., the Guilt and Shame Proneness Scale [GASP]; Cohen et al., 2011) present a series of guilt-inducing situations (e.g., revealing a friend's secret). Participants then report the likelihood that they would respond to the situation in a way that is indicative of guilt (e.g., exerting extra effort to keep secrets in the future). Although scenario-based scales are situation-specific, they have been criticized for their completely hypothetical structure. Guilt measured by scenario-based scales is predictive of empathic and ethical behavior (Tignor & Colvin, 2017, 2019) and is unrelated to poor outcomes like depressive symptoms (Kim et al., 2011). Such divergent findings across instruments suggest that guilt assessed with adjective-based scales is conceptually distinct from the guilt that is assessed with scenario-based scales. Tignor and Colvin (2019) found that guilt measured by an adjective-based scale was correlated with the propensity to experience guilt in daily life, whereas guilt assessed with a scenario-based scale was not. Although adjective-based scales exhibit better construct validity in this regard, their exclusion of guilt-inducing situations is a major drawback. A more practical measure would incorporate guilt's context-specificity, assess lived experiences of guilt, and avoid hypotheticals.

In parenting domains, guilt is typically associated with negative outcomes. Parental guilt concerning a broad variety of parenting practices, such as work-family balance, disciplinary practices, and emotion socialization, has been linked to depressive symptoms, anxiety, feelings of inadequacy, and child maladjustment (Borelli et al., 2017; Haslam et al., 2020; Seagram & Daniluk, 2002). Theoretical work posits that the primary driver of negative outcomes associated with guilt is an underlying belief about one's self as inadequate (Leach, 2017). In line with this notion, findings that guilt is largely maladaptive for parents may be due to the centrality of parenting to one's identity (Scarnier et al., 2009; Seagram & Daniluk, 2002).

Importantly, assessments of parental guilt have avoided the major drawbacks of more general adjective-based and scenario-based scales. Measures such as the Guilt About Parenting Scale (GAPS; Haslam et al., 2020) and the Pomona Work and Family Assessment (PWFA; Borelli et al., 2017) assess guilt in a context-specific manner, unlike adjective-based scales, and capture propensity for lived experiences of guilt rather than hypothetical feelings or responses, unlike scenario-based scales. Such assessments reveal that parental guilt is characterized by affective and behavioral tendencies that are classically indicative of guilt (e.g., motivation to reduce harm consistent with "approach-and-amend" themes), yet these tendencies are correlated with negative outcomes. These results are likely attributable to the non-hypothetical nature of parental guilt assessments, as well as the identity-centrality of parenthood.

Many studies of parental guilt regarding child food choices have been qualitative. In discussions of child feeding, child weight, and child eating habits, guilt and shame are core themes expressed by both mothers (Jackson et al., 2007; Johnson et al., 2011; Noble et al., 2005; Pocock et al., 2010) and fathers (Harris et al., 2020). Content analyses show that guilt arises from many aspects of child feeding, including factors within one's control, such as a desire to secure one's child's affection by providing unhealthy foods, and outside one's control, such as time and cost constraints (Jackson et al., 2007; Pescud & Pettigrew, 2014). Qualitative themes suggest that parents' food-related guilt is strongly associated with current unhealthy parental food choices.

Quantitative findings from studies of parents' food-related guilt are inconclusive, though there is a general tendency for guilt to be associated with desire or intentions to improve habits in the future which does not translate into such behavior. Persky and colleagues (2020) found that health-related guilt was *not* associated with healthier food choices for one's child, whereas Hollister and colleagues (2019) found guilt to be associated with feeling able to "take action" for one's child's health in the future. In another study, healthier food choice was associated with reductions in guilt about genetic obesity risk transmission (Persky et al., 2015). Finally, Hagerman and colleagues (2019) found that guilt was associated with stronger intentions to improve a child's diet in the future, but was *also* associated with unhealthier food choices and unhealthier food consumption by the parent at the time of observation, furthering the mixed nature of quantitative evidence.

Conflicting results are likely, in part, attributable to the measurement approaches used in past work. Parental guilt regarding child food choice is measured with inconsistent methods across studies, which are typically unvalidated and consist of only one or two items. For example, single-item measures were used by Hollister and colleagues

(2019) and Persky and colleagues (2020), whereas Hagerman and colleagues (2019) and Persky and colleagues (2015) each averaged two items. These items are often created *de novo* for use in a given application, and although they are reported as measuring the same construct, items differ in content from study to study.

Assessing guilt with one or two unvalidated items presents both conceptual and statistical problems. First, this approach is unlikely to successfully capture parents' propensity for lived experiences of food-related guilt. Many food-related situations may elicit guilt, including choices made during grocery shopping, eating out, and cooking (Ogle & Park, 2018; Pescud & Pettigrew, 2014); therefore, a more comprehensive set of items is needed to achieve validity and reliability. Indeed, other instruments that measure parents' feelings about child feeding (e.g., the Child Feeding Questionnaire [CFQ]; Birch et al., 2001) contain many items designed to span the range of situations that may elicit the construct under study. A measure that successfully captures parents' propensity to experience food choice-related guilt must be similarly comprehensive. Second, the reliability of a scale is strongly related to the number of items in that scale; Schmidt and Hunter (1996) note that a single-item measure is unlikely to yield reliability greater than 0.25. When the reliability of measurement instruments is reduced, observed relations among variables are weakened and replication becomes less likely (Muchinsky, 1996).

The goal of this study was to develop and evaluate a novel self-report instrument to measure parents' guilt about food choices for their children. The scale is intended for use among parents of 3-to-13-year-old children, as is typical among other parenting guilt and child feeding scales (e.g., Birch et al., 2001; Haslam et al., 2020). Given the broader goal of identifying links between parental guilt and food choices over time, this scale may be particularly useful for longitudinal work, but will also be appropriate for cross-sectional work. This tool will be broadly applicable across important health psychology and behavioral medicine contexts, including assessing the efficacy of obesity prevention interventions and elucidating the potential positive and/or negative consequences of food-related guilt for both parents and children.

Methods

Construct Definition

We began the scale development process by laying out a clear and comprehensive definition of our construct. To define parental guilt about food choices for the child, we draw upon previously formulated and well-established definitions of guilt (Kugler & Jones, 1992). Our definition of parental guilt about child food choices is as follows: a dysphoric state, affectively characterized by feelings of regret and remorse, behaviorally characterized by a desire to make reparation, *specifically* brought on by one's practices, habits, and decisions about the quality, quantity, and context of one's child's dietary intake.

Item Development

Item Content

Items were developed through an iterative process, beginning with a literature review pertaining to guilt, shame, parental affect, and measurement considerations for each. We structured items such that the identified drawbacks of scenario- and adjective-based scales would be avoided. Items were designed to capture experiences of guilt, within the specific context of child food choices, while avoiding hypotheticals. Each item was structured as follows: "When I think about [aspect of food choice for the child], I feel [response]."

By drawing upon experiences cited by parents in qualitative work and holding a focus-group style discussion with local parents and social scientists, we compiled a pool of salient aspects of child food choice with the potential to elicit guilt. These were paired with affective and behavioral responses indicative of guilt, also based on qualitative work and theory. Affective responses consisted of feeling regret or remorse (e.g., "I feel regretful," "I feel guilty"), whereas behavioral responses were characterized by wanting to fix the situation or do better in the future (e.g., "I feel like I want to do better for my child," "I want to fix it"). We avoided responses that would introduce confounds with shame (e.g., "I feel like a bad person").

Having created an initial pool of 18 items, we sought content validity feedback from six experts across domains of guilt and child feeding. Experts rated each item's quality on a 1–7 scale and were invited to provide open-ended comments. Items that received negative feedback were revised or removed. For example, a removed item read: "When I think about the food decisions I usually make for my child, I want to apologize to him/her." The desire to apologize was a common manifestation of parental guilt identified in the qualitative literature; however, multiple content experts noted that children would be unlikely to understand an apology in this regard, and therefore the item would likely yield noise. New items were created based on experts' recommendations. Following this, the new pool still contained 18 total items.

Defining the Outcome Space

Given the novel structure of our items, we took an empirical approach to determining the outcome space (i.e., the response set that accompanies each item). A small set of pilot data (N = 13) from parents of 3-to-13-year-olds was collected through Amazon Mechanical Turk (mTurk). Participants viewed each item followed by two potential response sets (Option A: 1 = *not at all true*, 5 = *extremely true*; Option B: 1 = *strongly disagree*, 5 = *strongly agree*). Participants indicated which response set was the most logical match for the preceding statement. Across all items, Option A was chosen 58% of the time and was chosen by most parents for all but three items. Therefore, our outcome space was defined as 1 = *not at all true*, 2 = *somewhat true*, 3 = *moderately true*, 4 = *very true*, 5 = *extremely true*, and the resulting data were treated as ordinal.

Preliminary Item Refinement

Items were refined based on feedback from a larger sample of parents of 3-to-13-year-olds (N = 118, 70 mothers, 48 fathers) recruited from a database of local parents interested in research. The goal was to preliminarily assess item performance with descriptive statistics, including response distributions and mean ratings of clarity and relevance. Thus, we did not recruit a sample as large as that required for Item Response Theory (IRT) or related analyses. In an online survey, participants responded to each item and then rated each item on relevance to their own life and clarity. Participants were encouraged to provide optional open-ended feedback on each item. See Supplemental Tables 1 and 2 for demographics and results; all Supplemental Materials are available at https://osf.io/9mvhp/.

Four items were removed due to serious floor effects and relevance ratings below the midpoint. In open-ended feedback, many parents noted that the aspects of child feeding captured by those items were not salient in real life. Item-total correlations were also low. Other items elicited negative feedback due to unintuitive wording; these were reworded and flagged for the next round of evaluation. Following this stage, there were 14 items.

Item Reduction

In this phase we collected data from 513 parents of 3-to-13-year-olds (306 mothers, 207 fathers) recruited through mTurk. See Supplemental Table 3 for demographics. Participants responded to the 14-item scale. Responses were analyzed using exploratory factor analysis (EFA) and a Graded Response Model (GRM) from the IRT framework, described further in the next section. The purpose of these analyses was to investigate the dimensionality of parental food-related guilt as measured by our items, and to formally evaluate item quality.

EFA indicated that the 14-item scale was multidimensional. The best-fitting solutions had 2–4 factors; however, the eigenvalues corresponding to Factors 2–4 were small relative to Factor 1, suggesting a dominant general factor (Supplemental Figure 1). Therefore, we examined the possibility of a bifactor structure with bifactor EFA. Results indicated the presence of one general factor and three subfactors. Groups of items associated with each subfactor tended to be adjacent and/or contained similar content, indicating redundancy (Supplemental Table 4). Additionally, the explained common variance statistic from the bifactor model was 0.811, indicating that 81.1% of the total variance explained by Factors 1–4 was attributable to Factor 1 (Reise, 2012). Most item variance was explained by a single factor, suggesting that this model could be reduced to a unidimensional model by removing redundant items.

If there is a dominant general factor, it is appropriate to fit a GRM that assumes a unidimensional latent structure (Cho et al., 2015). We fit this model to further assess item quality. Item quality was defined by goodness-of-fit to the data and item discrimination, each of which can be assessed within the IRT framework. Item discrimination is an item's ability to differentiate between individuals with different latent levels of the construct being measured.

Items were removed based on multiple criteria. Within each subfactor identified by the bifactor EFA, one item was retained to eliminate redundancy and achieve unidimensionality. The best-performing items were identified by relevance ratings from the first round of data collection, substantive justification, and item quality as indicated by the GRM (i.e., discrimination and goodness-of-fit to the data). Based on these criteria, 7 items were removed; see Supplemental Table 5 for details. The result was a 7-item scale. Before evaluating the final scale on a new sample, we used IRT results to compare the performance of the 14- and 7-item versions. Across latent guilt scores from about −2 to 2, the 7-item version yielded reliability that has been deemed acceptable for other self-report measures used as research tools rather than for individual assessment (e.g., Hornsby et al., 2021; see Supplemental Figures 2 and 3). As the intended use of the PFCG is not individual assessment or diagnosis, we proceeded with the 7-item scale.

Regarding sample size in this and the following phase of data collection, we recruited larger samples because IRT analyses were conducted. No definitive sample size guidelines exist for item response models. It has been proposed that N be proportional to the number of parameters in the model, though the recommended ratio varies (e.g., a minimum of 2 or 5 persons per estimated parameter; de Ayala, 2009). In this phase, the GRM had 70 estimated parameters; 14 items = 14 discriminations + (14 × 4 thresholds). In the next phase, the GRM had 35 estimated parameters; 7 items = 7 discriminations + (7 × 4 thresholds). Following the recommendation of a 1:5 ratio of model parameters to persons, minimum N would be 350 and 175 persons, respectively, in these phases. Because these are rough guidelines, we opted for larger samples to ensure adequate precision of estimates.

Final Evaluation of the Parental Food Choice Guilt Scale (PFCG)

Participants and Recruitment

We formally evaluated all psychometric properties of the 7-item PFCG on a final sample of parents. Recruitment took place via CloudResearch, a crowdsourcing platform that overlays mTurk. We incorporated multiple types of "validity indicators" throughout our online survey, including short attention checks and open-ended questions with which we could detect answers indicative of unreliable data (e.g., cheating, bots). Participants who failed these checks (N = 13) were removed, yielding N = 294.

Participants had at least one child between 3 and 13 years of age. Although feeding behavior certainly changes across this child age range, past work indicates that parents' thoughts and feelings about their food choices can be measured comparably across this range. For example, the Child Feeding Questionnaire (CFQ; Birch et al., 2001) is designed for use among parents of children aged 2–11, and its psychometric properties are well established; the Meals in our Household Questionnaire (Anderson et al., 2012), which assesses mealtime family environment, was validated by parents of children aged 3–11; and the Family Mealtime Goals Questionnaire (Snuggs et al., 2019) was developed for use among parents of children aged 1–16 years. In a systematic review, Vaughn and colleagues (2013) summarize other instruments designed to measure parents' perceptions of their feeding practices and the at-home food environment, many of which were validated on parents of children spanning a 10-year age range or more. To our knowledge, no parental guilt- or feeding-related scales have been shown to behave differently according to child age within similar ranges.

After eligibility screening, if the participant had more than one child between the ages of 3 and 13, they were instructed to only think about the oldest child in that age range as they answered subsequent questions. Participants were compensated via the mTurk payment system. All waves of data collection were approved by the Institutional Review Board of the National Human Genome Research Institute. See Table 1 for demographics.

Measures

Parents first completed the 7-item PFCG with item order randomized. Next, we measured guilt- and shame-proneness with the GASP, a scenario-based scale that includes four subscales: Guilt Negative Behavior Evaluation, Guilt Repair, Shame Negative Self-Evaluation, and Shame Withdrawal (Cohen et al., 2011). Then, parents responded to the Guilt About Parenting Scale (GAPS; Haslam et al., 2020), which assesses general parenting-related guilt. Self-efficacy about child feeding was measured with an adaptation of the Parenting Sense of Competence Scale (Johnston & Mash, 1989). Parents completed the Concern About Child Weight and Perceived Feeding Responsibility subscales of the CFQ (Birch et al., 2001).

Eight face-valid items designed to assess guilt about various aspects of child health, which have been used in previous studies but have not been psychometrically validated, were included next. These items began with "I feel guilty…" followed by "about my child's eating habits," "about my child's physical activity habits," "about the genetic risk for obesity that I may have passed down to my child," "that our home environment could increase my child's risk for obesity," "about my child's sleep habits," and "about my child's overall risk for obesity later in life." Two items began with "If my child were to develop obesity later in life, I would feel…" followed by guilty and regret. All were 7-point Likert-type scales (1 = *strongly disagree*, 7 = *strongly agree*).

Parents' perceptions of their child feeding practices were assessed with the Modeling subscale of the Comprehensive Feeding Practices Questionnaire (CFPQ; Musher-Eizenman & Holub, 2007). We measured child eating habits with the Food Fussiness subscale of the Child Eating Behavior Questionnaire (CEBQ; Wardle et al., 2001) and two face-valid items created for this study ("How healthy do you think your child's diet is?" 1 = *Very unhealthy*, 7 = *Very healthy*, and "On a typical day, how close does your child get to eating the recommended number of fruits and vegetables?" 1 = *Not close at all*, 7 = *Very close/meets recommendation*).

Next, we included a set of questionnaires for a different study concerning general beliefs about heritability and perspective-taking; these served as distractors for the current study. Following this, we assessed guilt-proneness with the PANAS-X (Watson & Clark, 1994) as well as guilt- and shame-proneness with the PFQ (Harder & Lewis, 1987), both adjective-based scales. Finally, parents reported demographic information for themselves and their children.

Analytic Strategy

First, we used EFA to assess the dimensionality of food-related guilt as measured by the PFCG items. In determining the number of dimensions to extract, parallel analysis is considered the most accurate method (Horn, 1965; Humphreys & Montanelli, 1975; Velicer et al., 2000). Parallel analysis is a simulation method that compares observed eigenvalues with those obtained from uncorrelated normal variables. It is recommended that a dimension be retained if an eigenvalue from the observed data is larger than the corresponding eigenvalue from the random

Table 1. Participant demographic characteristics (*N* = 294)

Characteristic	Parent *N* (%) or *M* (*SD*)	Child *N* (%) or *M* (*SD*)
Age (years)	37.62 (6.73)	8.42 (3.36)
Gender		
Female	194 (66.0%)	136 (46.6%)
Male	99 (33.7%)	152 (51.7%)
Nonbinary/third gender	1 (0.3%)	–
Weight Status		
Very overweight	31 (10.5%)	1 (0.3%)
Overweight	127 (43.2%)	10 (3.4%)
A little overweight	–	40 (13.6%)
About right	121 (41.2%)	211 (71.8%)
A little underweight	–	26 (8.8%)
Underweight	10 (3.4%)	3 (1.0%)
Very underweight	1 (0.3%)	0 (0%)
Marital Status		
Married	197 (67.0%)	–
Cohabiting	35 (11.9%)	–
Single	32 (10.9%)	–
Divorced/separated	24 (8.2%)	–
Widowed	2 (0.7%)	–
Education		
Doctoral degree	11 (3.7%)	–
Master's degree	45 (15.3%)	–
Bachelor's degree	106 (36.1%)	–
Trade school	17 (5.8%)	–
Some college	72 (24.5%)	–
High school/GED	37 (12.6%)	–
Some high school	2 (0.7%)	–
Race		
White	243 (82.7%)	–
Black/African-American	23 (7.8%)	–
Multiracial	10 (3.4%)	–
Asian	7 (2.4%)	–
Other	3 (1.0%)	–
Native American/Alaska Native	3 (1.0%)	–
Native Hawaiian/Pacific Islander	1 (0.3%)	–

data. A strength of this method is that it permits clear conclusions about how many dimensions to extract. Because the PFCG has ordinal responses, we conducted parallel analysis with polychoric correlations (Cho et al., 2009). We also examined descriptives for each item, including response spreads (i.e., histograms), means, standard deviations, and item-total correlations.

We evaluated item and scale quality using IRT analyses. Compared to analyses rooted in Classical Test Theory (CTT), IRT analyses present many important benefits. IRT yields item-level information whereas CTT methods evaluate the measure as a whole. IRT also provides precise estimates of individuals' *latent* rather than *observed* scores. IRT allows for nonconstant standard error of measurement across different levels of the latent trait, which here allows us to examine the latent levels of parental guilt at which the scale is most accurate. Finally, IRT permits generalization of results to larger populations rather than limitation to a specific sample, which is a drawback of CTT (Reise et al., 2005).

A Graded Response Model (GRM) with linear parameterization was used to investigate item and scale behavior. The GRM is the most frequently used item response model in clinical and psychological science because it yields intuitive estimates that are particularly appropriate for Likert-type scales characterized by ordered categorical responses (Samejima, 2016). In the GRM, two kinds of item parameters are estimated: discriminations and thresholds.

Discrimination estimates reflect each item's ability to differentiate among individuals along the latent continuum, or in other words, individuals with various levels of the latent construct being measured. Discrimination estimates should be positive and large; in general, < 0.64 is considered low, 0.65–1.34 moderate, and > 1.35 is considered large (Baker, 2001). Four thresholds were estimated for each item because the response set had five categories. In the GRM used here, thresholds were cumulative. An item's first threshold is the point along the latent trait continuum at which there is equal probability of endorsing the first response option vs. all response options above; the second threshold is the point at which there is equal probability of endorsing the first or second response option vs. the three response options above; and so on. Thresholds are evaluated by whether they are well-separated across the latent trait continuum; if so, this indicates that items and their response anchors behave as intended (Wilson, 2004). Each item's fit to the data was assessed with the S-χ^2 test; the null hypothesis is that the item fits well to the data, and high p-values indicate the discrepancy between the model and the data is non-significant (Kang & Chen, 2008).

To evaluate the scale as a whole, we examined measures of model-data fit including the C_2 measure (Cai & Monroe, 2014), RMSEA and its 95% confidence interval, SRMR, TLI, and CFI. The C_2 statistic was developed primarily for polytomous response models where there are not sufficient degrees of freedom to compute M_2, which was the case here. We also examined the test information function (TIF), which is a measure of scale performance unique to IRT, and presents the amount of information (i.e., certainty about latent scores) that the scale provides across different levels of the latent construct. Test information is inversely related to the standard error of latent scores; therefore, higher test information indicates better measurement fidelity.

Construct validity is the extent to which an instrument measures the construct that it is intended to measure. Evidence for construct validity can be accrued incrementally through a variety of methods (Cronbach & Meehl, 1955), including demonstrating convergent validity, discriminant validity, and establishing a lack of differential item functioning (DIF). We assessed convergent and discriminant validity via correlations among the PFCG and other relevant self-report measures. We tested for DIF with respect to parent gender, parent weight status, child weight status, and parental education. DIF analysis is unique to the IRT framework and tests whether items yield different parameter estimates across groups, thereby assessing whether the item measures the same latent construct across groups. We used an ordinal logistic regression model with latent scores as a matching criterion. DIF was detected by comparing three nested ordinal logistic regression models with likelihood ratio tests, as described by Crane and colleagues (2006). Across the models, the difference in deviance statistics is compared to a χ^2 distribution with $df = 1$. A Type I error rate of $\alpha = .01$ was chosen. For each item, we tested for *uniform* DIF, where the difference across groups is constant, and *nonuniform* DIF, where the difference across groups varies by latent trait level. McFadden's pseudo R^2 measure, which reflects a proportional reduction in the deviance statistic, was chosen for a DIF effect size measure. Zumbo's (1999) guidelines for classifying DIF based on the pseudo R^2 statistic are: negligible (< 0.13), moderate (0.13–0.26), and large (> 0.26).

Reliability was quantified with Cronbach's α and the marginal IRT reliability coefficient; this coefficient is interpreted as the proportion of observed score variance that is attributable to true individual differences rather than measurement error (Wright & Masters, 1981). Analyses were conducted in R using the packages *random.polychor.pa* (Presaghi et al., 2019), *mirt* (Chalmers, 2012), and *lordif* (Choi et al., 2011).

Results

Exploratory Factor Analysis

EFA results strongly supported a one-factor solution (Figure 1). The first eigenvalue was large, whereas all others were less than 1 (Factor 1 = 4.52, Factor 2 = 0.49, Factor 3 = 0.46). Accordingly, parallel analysis indicated that the first factor should be extracted whereas the subsequent factors should not. Overall, results pointed to a unidimensional latent structure.

Item Descriptive Information

Response spreads indicated there was a sufficient number of respondents within each response category for each item (de Ayala, 2009). Ceiling and floor effects did not arise. All items yielded strong positive item-total correlations ($0.78 < r < 0.84$, all ps < .01). Correlations > 0.40 preliminarily indicate good discriminatory ability (Ebel, 1965). See Table 2.

Item Response Model Comparisons

We began by examining whether the GRM was the most appropriate model over others that are suitable for polytomous data. We compared models by examining model selection criteria. Compared to the Rating Scale Model (AIC = 5,263.75, BIC = 5,304.27) and the Partial Credit

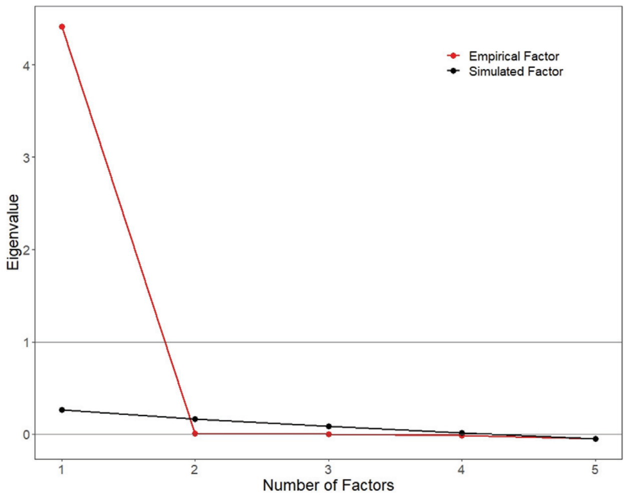

Figure 1. Eigenvalues obtained from parallel analysis using polychoric correlations on the 7-item PFCG

Table 2. Item descriptives

	Response frequency (*N*)						
Item	1	2	3	4	5	Item location (Mean)	Item-total correlation
1	56	92	66	53	27	2.67	0.80
2	53	90	59	69	23	2.72	0.79
3	56	92	68	49	29	2.67	0.78
4	64	91	64	55	20	2.58	0.80
5	50	90	65	56	33	2.77	0.84
6	68	86	67	48	25	2.58	0.80
7	54	96	63	63	18	2.64	0.81

Model (AIC = 5,252.04, BIC = 5,351.50), the GRM yielded the best AIC and similar BIC (AIC = 5,248.86, BIC = 5,377.79). In the presence of inconclusive model selection criteria, we noted that from a theoretical perspective, the GRM is the preferred model for Likert-type response scales like the PFCG. Therefore, we proceeded with the GRM for further analysis.

Item Response Model Estimates and Fit

For the model-data fit of the GRM, we observed $C_2 = 11.23$, $df = 14$, $p = .67$. Here, the null hypothesis is that the model fits exactly in the population; therefore, a nonsignificant *p*-value indicates a satisfactory model-data fit. Additionally, we observed RMSEA = 0 [0, 0.045], SRMR = 0.023, TLI = 1, and CFI = 1, all of which indicated good model-data fit.

The TIF for the 7-item PFCG is shown in Supplemental Figure 4. Test information was highest along the middle range of latent trait scores, ranging from −2 to 2. The PFCG is best able to differentiate among individuals with mild-to-moderate levels of guilt.

Item discriminations ranged from 2.121 to 2.963, which are large values (Baker, 2001; see Table 3). Threshold estimates suggested that our items successfully span the range of the latent trait continuum; the lowest threshold estimates ranged from −3.206 to −2.199, and the highest ranged from 3.554 to 4.715. Thresholds ascended monotonically and were well-separated in terms of location, indicating that items and response anchors behave well. Standard errors (*SE*s) of all estimates were acceptable (< 0.43; Wilson, 2004). $S\text{-}\chi^2$ indices showed that all items fit well to the data, as all *p*-values were > .05 (see Supplemental Table 6).

Latent guilt estimates ranged from −2.051 to 2.569 ($M = 0.00$, $SD = 0.96$). *SE*s ranged from 0.243 to 0.489, indicating acceptable precision (Wilson, 2004).

Scale Reliability

Small *SE*s of latent trait scores were our first indicator of good reliability (Wilson, 2004). Marginal IRT reliability was 0.909. In addition, Cronbach's α = .90.

Convergent and Discriminant Validity

The PFCG was negatively correlated with parents' feeding self-efficacy ($r = -0.29$, $p < .001$), parent-reported healthfulness of the child's diet ($r = -0.35$, $p < .001$), and child fruit and vegetable intake ($r = -0.25$, $p < .001$). Positive correlations arose with child food fussiness ($r = 0.29$, $p < .001$) and concern about child weight ($r = 0.48$, $p < .001$). The PFCG was uncorrelated with the guilt-related subscales of the GASP (*r*s = 0.04, 0.02, *p*s = .50, .74). In contrast, the PFCG was positively correlated with guilt subscales of the PANAS-X ($r = 0.27$, $p < .001$) and the PFQ ($r = 0.23$, $p < .001$), as well as the shame subscale of the PFQ ($r = 0.22$, $p < .001$) and the shame-withdrawal subscale of the GASP ($r = 0.24$, $p < .001$). The PFCG was also positively correlated with the GAPS ($r = 0.34$, $p < .001$), which itself displayed positive correlations with shame and guilt subscales of the PANAS-X and PFQ. The PFCG was positively correlated with face-valid single-item assessments of guilt about various aspects of child health, especially those concerning the child's eating habits ($r = 0.60$, $p < .001$), the child's risk for obesity later in life ($r = 0.54$, $p < .001$), and the home environment ($r = 0.51$, $p < .001$). See Figure 2.

Differential Item Functioning (DIF)

The grouping variables examined were parent gender, reported weight status of the parent (overweight or very overweight vs. all other categories), reported weight status of the child (a little overweight, overweight, or very overweight vs. all other categories), and parent education

Table 3. Item parameter estimates (*SE*s) of the Graded Response Model

Item	α (*SE*)	δ_1 (*SE*)	δ_2 (*SE*)	δ_3 (*SE*)	δ_4 (*SE*)
1	2.496 (0.251)	−2.578 (0.280)	0.075 (0.215)	1.852 (0.250)	4.050 (0.367)
2	2.343 (0.238)	−2.635 (0.275)	−0.113 (0.205)	1.386 (0.226)	4.138 (0.371)
3	2.121 (0.216)	−2.391 (0.249)	−0.056 (0.193)	1.653 (0.224)	3.554 (0.319)
4	2.340 (0.237)	−2.294 (0.255)	0.103 (0.207)	1.898 (0.246)	4.389 (0.390)
5	2.963 (0.307)	−3.206 (0.343)	−0.200 (0.242)	1.744 (0.276)	4.177 (0.409)
6	2.380 (0.242)	−2.199 (0.254)	0.090 (0.209)	1.971 (0.250)	4.014 (0.360)
7	2.490 (0.251)	−2.730 (0.288)	0.143 (0.217)	1.891 (0.252)	4.715 (0.422)

Note. α = item discrimination; δ_1 = item's first threshold; δ_2 = item's second threshold; etc.

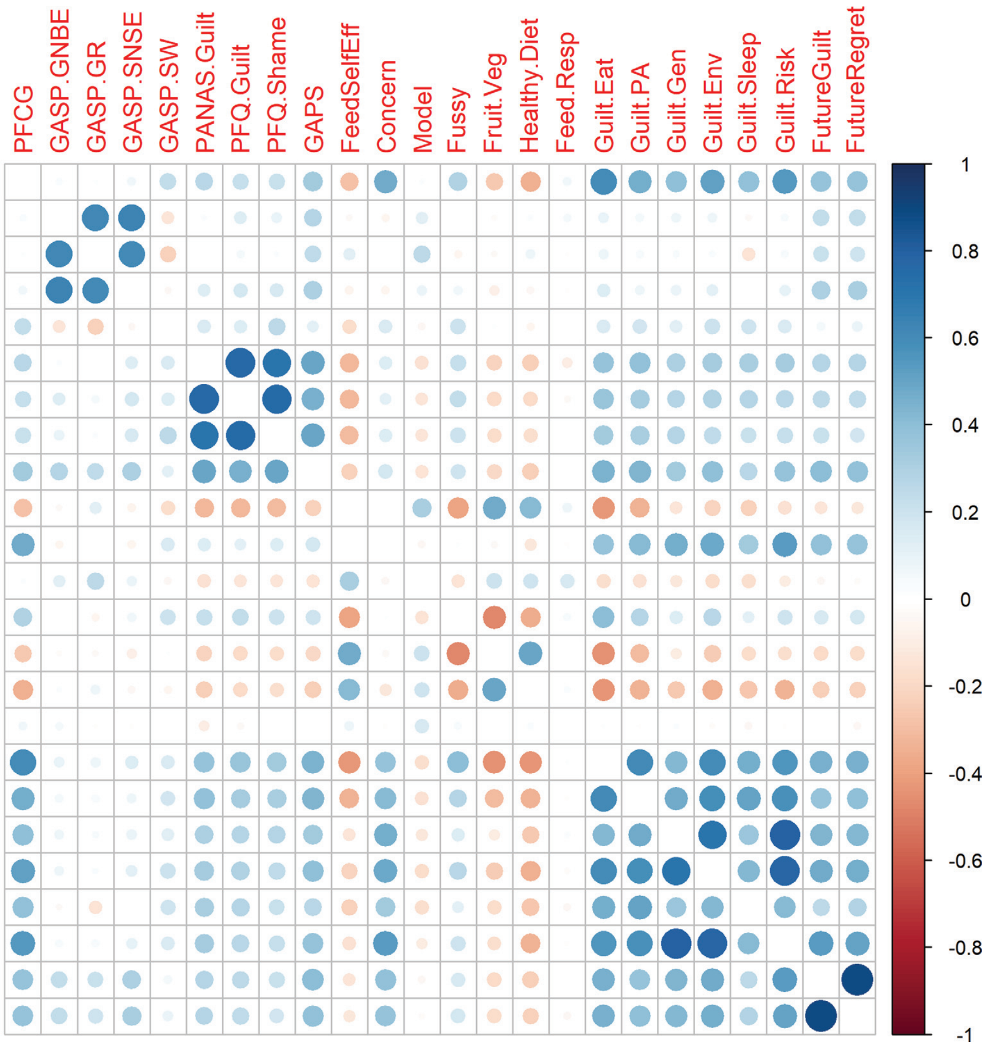

Figure 2. Intercorrelations among the PFCG and related self-report measures. PFCG = Parental Food Choice Guilt Scale; GASP.GNBE = GASP Guilt-Negative Behavior Evaluation; GASP.GR = GASP Guilt-Repair; GASP.SNSE = GASP Shame-Negative Self Evaluation; GASP.SW = GASP Shame-Withdrawal; GAPS = Guilt About Parenting Scale; FeedSelfEff = feeding self-efficacy; Concern = CFQ Concern about Child Weight; Model = CFPQ Healthy Food Modeling; Fussy = CEBQ Child Food Fussiness; Guilt.Eat, Guilt.PA, Guilt.Gen, Guilt.Env, Guilt.Sleep, FutureGuilt, and FutureRegret = face-valid guilt items.

(college graduate vs. not) which was our closest proxy to socioeconomic status. Evidence suggests that fathers play similarly important roles as mothers in child feeding (Garfield & Isacco, 2012; Khandpur et al., 2014), so we hypothesized no DIF across parent gender. We expected that DIF may emerge across parental and/or child weight status, such that it would be "easier" to endorse guilt among parents with overweight, or parents of children with

Table 4. The Parental Food Choice Guilt Scale

Prompt

Below are some statements describing parents' feelings about the way they feed their children. Please indicate the extent to which each statement is true.

Response set

1 = *Not at all true*	2 = *Somewhat true*	3 = *Moderately true*	4 = *Very true*	5 = *Extremely true*

Items

1. When I think about the foods I usually buy for my child at the store, I feel like I should make better choices.
2. When I think about the times I've "given in" to my child's requests for junk food, I feel guilty.
3. When I think about the types of foods I usually let my child order when we eat out, I feel regretful.
4. When I think about the sugar content of the food my child typically eats, I feel like I want to do better for my child.
5. When I think about the types of food my child usually eats, I want to fix it.
6. When I think about the food decisions I usually make for my child, I feel like I want to do better.
7. When I think about the times I've fed my child unhealthy processed foods, I feel regretful.

Note. Randomize item order of appearance if possible.

overweight (Conradt et al., 2008). We expected items involving financial barriers to healthy feeding would be "easier" to endorse among noncollege-educated parents, as past work has shown that such barriers may contribute to unhealthier food choices and related guilt (Pescud & Pettigrew, 2014).

For all items, neither uniform nor nonuniform DIF was identified by any grouping variable. All *p*-values were > .01 and the pseudo R^2 statistic did not exceed .005 for any test, indicating extremely small effect sizes (see Supplemental Table 7). Results supported the construct validity of the PFCG, such that each item measures the same latent construct and behaves comparably regardless of parent gender, education, or parent/child weight status.

Discussion

The goal of this study was to develop and validate the PFCG, a novel measure of parental guilt about child food choices. The validated 7-item PFCG is shown in Table 4. The PFCG behaves very well psychometrically; each item displays high discriminatory power and thresholds that spread across the latent trait continuum. The PFCG yields the most precise latent trait estimates along the mid-range of the continuum, indicating a good match to the levels of feeding-related guilt that are most relevant for research and practice. The PFCG is highly reliable and does not systematically differ across important demographic groups.

EFA results, alongside strong evidence for goodness-of-fit of the unidimensional GRM, strongly supported unidimensionality. Although more general measures of guilt-proneness sometimes display multidimensional structures (e.g., distinct factors for affective vs. behavioral manifestations of guilt), PFCG items did not load onto different factors according to whether they assessed behavioral tendencies or affect. This result mimics other context-specific measures of guilt like the Guilt About Parenting Scale (GAPS; Haslam et al., 2020). Our results further support the notion that within parenting contexts, guilt can be assessed according to a single underlying factor. However, it is important to note that we began with a small item pool. An initial pool of 18 items is smaller than many IRT applications that begin with dozens or hundreds of items. The construct measured by the PFCG is narrow, and the qualitative literature pertaining to this topic is similarly narrow. These 18 items largely exhausted identified themes from this literature and scenarios generated by focus group parents, so it was difficult to generate a larger pool of items such as those developed for broad constructs like depression. This small initial item pool may have contributed to the ultimate unidimensionality of the PFCG.

Importantly, there are many aspects of child feeding that likely elicit parental guilt, but that the PFCG does not capture. "Child feeding practices" are often defined broadly to include a variety of behaviors, such as monitoring the child's food intake and modeling healthy eating (e.g., Birch et al., 2001). Most PFCG items assess guilt related specifically to the nutritional content of foods that the parent chooses for the child, or that the parent allows the child to choose. Our initial item pool contained items that captured other aspects of child feeding; for example, two items involved using food to manage child behavior (e.g., "When I think about the times I've used food as a reward for my child, I want to go back and change what I've done"), and two items addressed effort toward meal preparation (e.g., "When I think about the times I've fed my child unhealthy food because it's easier, I feel regretful"). However, all these items were dropped throughout the refinement stages. Most were dropped either due to content experts' arguments for lack of generality or low relevance ratings from parents themselves. Conceptually, it makes sense that guilt related to these other aspects of feeding would be less universal, as they intersect with parenting domains where there is high variability in experience and practice. These items may have also felt more difficult to endorse because they were qualitatively different from the larger pool of questions, most of which were narrowly focused on food choice. Developing assessments of guilt regarding these broader aspects of child feeding will be a useful avenue for future research; such an undertaking would almost certainly yield a multidimensional scale, and it will be important to establish relevance and construct validity for each of these dimensions.

The PFCG was negatively correlated with the healthfulness of the child's diet and positively correlated with the child's food fussiness. These findings align with past work linking guilt with poorer child eating and feeding habits at the moment (Hagerman et al., 2019; Johnson et al., 2011; Noble et al., 2005; Pescud & Pettigrew, 2014; Pocock et al., 2010), supporting the idea that, rather than driving adaptive behavior, parental food-related guilt is a correlate of relatively unhealthy parent and child behaviors. Our study, alongside most prior work, was cross-sectional, so the directionality of these relationships is unclear. Unhealthy child eating may elicit parental guilt, or greater parental guilt may drive poorer health behaviors. These relations are likely bidirectional and unfold over time, and further work is needed to elucidate their precise nature.

Providing further evidence of construct validity, the PFCG was positively correlated with guilt measured by the PFQ and the PANAS-X, both adjective-based scales, and uncorrelated with guilt subscales of the GASP, a scenario-based scale. Guilt assessed with adjective-based scales is associated with the propensity to experience guilt in daily life, whereas guilt assessed with hypothetical scenario-based scales is not (Tignor & Colvin, 2019). Observed correlations with the PFQ and PANAS-X confirm that our scale captures parents' lived experiences of guilt. Avoiding the decontextualized nature of adjective-based measures, the PFCG assesses guilt responses to specific aspects of food choice. Avoiding the hypothetical nature of scenario-based measures, items capture actual – rather than imagined – guilt responses.

The PFCG was positively correlated with general parenting guilt assessed with the GAPS and negatively correlated with feeding self-efficacy. The GAPS captures guilt related to time and work constraints (e.g., "I feel bad if I am not at home or with my family") and discipline (e.g., "I feel bad if I am inconsistent in parenting or disciplining my child"). Indeed, parents' food-related guilt is often intertwined with time and cost constraints, as well as uncertainty about appropriate levels of control to implement in monitoring children's diet (Noble et al., 2005; Pescud & Pettigrew, 2014). Our results with respect to self-efficacy are consistent with qualitative findings linking food-related guilt to less parental knowledge about nutrition (Ogle & Park, 2018). However, it is also important to note that guilt and shame may be particularly difficult to disentangle among parents (Scarnier et al., 2009). It is possible that responses to our scale were driven by a general belief about inadequacy as a parent (Leach, 2017), which would explain observed relations with general parenting guilt and feeding self-efficacy.

The PFCG was positively correlated with single-item measures of parents' health behavior-related guilt. For example, a correlation of 0.60 was observed between the PFCG and "I feel guilty about my child's eating habits." While this result supports the construct validity of the PFCG, it also demonstrates the inadequacy of a single-item measure. Squaring this correlation, we find that the PFCG shares only 36% of its variance with this item. Put together, PFCG items capture guilt responses to many aspects of parents' food choices, creating a more comprehensive and realistic measure of parents' felt guilt. Most of the variance in these responses is not captured by single-item measures of guilt, demonstrating the incremental validity of our scale.

Strengths of the study included iterative scale construction grounded in theoretical and empirical support and evaluation with IRT analyses. Our samples were evenly split by parent gender, enabling us to examine DIF across parent gender. To our knowledge, related measures have not been evaluated in this way, despite calls to examine father involvement in child health (Garfield et al., 2019). No items displayed DIF by parent gender, further confirming that mothers and fathers play similar roles in child feeding (Garfield & Isacco, 2012; Harris et al., 2020; Khandpur et al., 2014). Beyond behaviors enacted, our results suggest that mothers' and fathers' *thoughts and feelings* about child food choices can be measured comparably.

There were limitations to this study. Most of our samples were skewed toward White, co-parenting, well-educated parents. Had our samples been more diverse, particularly throughout the item refinement and reduction stages, different results may have emerged with respect to item relevance and performance. Generalizations of these results to other demographic groups, such as low-income parents, should be made with caution. Data were cross-sectional, inhibiting our ability to examine test-retest reliability or criterion validity across time. Additionally, the PFCG measures parents' health-related guilt only regarding food choices. Future work should explore parental guilt in other health domains such as genetics, obesity, and physical activity, and strive to comprehensively measure these constructs with valid and reliable approaches.

The PFCG demonstrates excellent psychometric properties and is suitable for use with parents of 3-to-13-year-old children. With access to a theoretically informed and validated measure of parents' guilt regarding food choices for their children, future work will be better able to elucidate guilt's role in shaping behavior. This will be an essential component of broader efforts targeting parents' preventive health behaviors for their children. Valid and reliable instruments will provide the necessary foundation for this important work.

References

Anderson, S. E., Must, A., Curtin, C., & Bandini, L. G. (2012). Meals in our household: Reliability and initial validation of a questionnaire to assess child mealtime behaviors and family mealtime environments. *Journal of the Academy of Nutrition and Dietetics, 112*(2), 276–284. https://doi.org/10.1016/j.jada. 2011.08.035

Baker, F. B. (2001). *The basics of item response theory*. ERIC.

Birch, L. L., & Fisher, J. O. (2000). Mothers' child-feeding practices influence daughters' eating and weight. *American Journal of Clinical Nutrition, 71*, 1054–1061. https://doi.org/10.1093/ajcn/71.5.1054

Birch, L. L., Fisher, J. O., Grimm-Thomas, K., Markey, C. N., Sawyer, R., & Johnson, S. L. (2001). Confirmatory factor analysis of the Child Feeding Questionnaire: A measure of parental attitudes, beliefs and practices about child feeding and obesity proneness. *Appetite, 36*(3), 201–210. https://doi.org/10.1006/appe.2001.0398

Borelli, J. L., Nelson, S. K., River, L. M., Birken, S. A., & Moss-Racusin, C. (2017). Gender differences in work-family guilt in parents of young children. *Sex Roles, 76*(5–6), 356–368. https://doi.org/10.1007/s11199-016-0579-0

Cai, L., & Monroe, S. (2014). *A new statistic for evaluating item response theory models for ordinal data. (CRESST Report 839)*. National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Cho, S.-J., Li, F., & Bandalos, D. (2009). Accuracy of the parallel analysis procedure with polychoric correlations. *Educational and Psychological Measurement, 69*(5), 748–759. https://doi.org/10.1177/0013164409332229

Cho, S.-J., Wilmer, J., Herzmann, G., McGugin, R. W., Fiset, D., Van Gulick, A. E., Ryan, K. F., & Gauthier, I. (2015). Item response theory analyses of the Cambridge Face Memory Test (CFMT). *Psychological Assessment, 27*(2), 552–566. https://doi.org/10.1037/pas0000068

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software, 39*(8), 1–30. https://doi.org/10.18637/jss.v039.i08

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP Scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology, 100*(5), 947–966. https://doi.org/10.1037/a0022641

Conradt, M., Dierk, J., Schlumberger, P., Rauh, E., Hebebrand, J., & Rief, W. (2008). Who copes well? Obesity-related coping and its associations with shame, guilt, and weight loss. *Journal of Clinical Psychology, 64*(10), 1129–1144. https://doi.org/10.1002/jclp.20501

Crane, P. K., Gibbons, L. E., Jolley, L., & van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. *Medical Care, 44*(11), S115–S123. https://www.jstor.org/stable/41219511

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957

de Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.

Ebel, R. L. (1965). Confidence weighting and test reliability. *Journal of Educational Measurement, 2*(1), 49–57. https://www.jstor.org/stable/1433833

Garfield, C. F., Fisher, S. D., Barretto, D., Rutsohn, J., & Isacco, A. (2019). Development and validation of a father involvement in health measure. *Psychology of Men & Masculinities, 20*(1), 148–160. https://doi.org/10.1037/men0000144

Garfield, C. F., & Isacco, A. J. III (2012). Urban fathers' involvement in their child's health and healthcare. *Psychology of Men & Masculinity, 13*(1), 32–48. https://doi.org/10.1037/a0025696

Hagerman, C. J., Ferrer, R. A., Klein, W. M. P., & Persky, S. (2019). Association of parental guilt with harmful versus healthful eating and feeding from a virtual reality buffet. *Health Psychology, 39*(3), 199–208. https://doi.org/10.1037/hea0000831

Harder, D. W., & Lewis, S. J. (1987). The assessment of shame and guilt. In J. N. Butcher & C. D. Spielberger (Eds.), *Advances in personality assessment* (Vol. 6, pp. 89–114). Erlbaum.

Harris, H. A., Jansen, E., & Rossi, T. (2020). "It's not worth the fight": Fathers' perceptions of family mealtime interactions, feeding practices and child eating behaviours. *Appetite, 150*, Article 104642. https://doi.org/10.1016/j.appet.2020.104642

Haslam, D., Filus, A., & Finch, J. (2020). The Guilt about Parenting Scale (GAPS): Development and initial validation of a self-report measure of parenting guilt, and the relationship between parenting guilt and work and family variables. *Journal of Child and Family Studies, 29*(3), 880–894. https://doi.org/10.1007/s10826-019-01565-8

Hofmann, W., & Fisher, R. R. (2012). How guilt and pride shape subsequent self-control. *Social Psychological and Personality Science, 3*(6), 682–690. https://doi.org/10.1177/1948550611435136

Hollister, B. M., Yaremych, H. E., Goldring, M. R., & Persky, S. (2019). Mothers' and fathers' cognitive and affective responses to epigenetics concepts. *Environmental Epigenetics, 5*(4), 1–10. https://doi.org/10.1093/eep/dvz021

Horn, J. L. (1965). A rationale and test for the number of components from the matrix of partial correlations. *Psychometrika, 30*(2), 179–185. https://doi.org/10.1007/BF02289447

Hornsby, B. W., Camarata, S., Cho, S. J., Davis, H., McGarrigle, R., & Bess, F. H. (2021). Development and validation of the Vanderbilt Fatigue Scale for Adults (VFS-A). *Psychological Assessment, 33*(8), 777–788. https://doi.org/10.1037/pas0001021

Humphreys, L. G., & Montanelli, R. G. Jr. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research, 10*(2), 193–205. https://doi.org/10.1207/s15327906mbr1002_5

Jackson, D., Wilkes, L., & McDonald, G. (2007). "If I was in my daughter's body I'd be feeling devastated": Women's experiences of mothering an overweight or obese child. *Journal of Child Health Care, 11*(1), 29–39. https://doi.org/10.1177/1367493507073059

Johnson, C. M., Sharkey, J. R., Dean, W. R., McIntosh, W. A., & Kubena, K. S. (2011). It's who I am and what we eat. Mothers' food-related identities in family food choice. *Appetite, 57*(1), 220–228. https://doi.org/10.1016/j.appet.2011.04.025

Johnston, C., & Mash, E. J. (1989). A measure of parenting satisfaction and efficacy. *Journal of Clinical Child Psychology, 18*(2), 167–175. https://doi.org/10.1207/s15374424jccp1802_8

Kang, T., & Chen, T. T. (2008). Performance of the generalized $S\text{-}\chi^2$ item fit index for polytomous IRT models. *Journal of Educational Measurement, 45*(4), 391–406. https://doi.org/10.1111/j.1745-3984.2008.00071.x

Khandpur, N., Blaine, R. E., Fisher, J. O., & Davison, K. K. (2014). Fathers' child feeding practices: A review of the evidence. *Appetite, 78*, 110–121. https://doi.org/10.1016/j.appet.2014.03.015

Kim, S., Thibodeau, R., & Jorgensen, R. S. (2011). Shame, guilt, and depressive symptoms: A meta-analytic review. *Psychological Bulletin, 137*(1), 68–96. https://doi.org/10.1037/a0021466

Kugler, K., & Jones, W. H. (1992). On conceptualizing and assessing guilt. *Journal of Personality and Social Psychology, 62*(2), 318–327. https://doi.org/10.1037/0022-3514.62.2.318

Leach, C. W. (2017). Understanding shame and guilt. In L. Woodyat, E. L. Worthington, M. Wenzel, & B. J. Griffin (Eds.), *Handbook of the psychology of self-forgiveness* (pp. 17–28). Springer.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement, 56*(1), 63–75. https://doi.org/10.1177/0013164496056001004

Musher-Eizenman, D., & Holub, S. (2007). Comprehensive feeding practices questionnaire: Validation of a new measure of parental feeding practices. *Journal of Pediatric Psychology, 32*(8), 960–972. https://doi.org/10.1093/jpepsy/jsm037

Noble, G., Jones, S. C., & McVie, D. (2005). Motivational factors in the food buying behaviour of parents of pre-school age children: A projective technique study. In S. Purchase (Ed.), *Proceedings for the ANZMAC Conference, University of Western Australia* (pp. 164–170). https://ro.uow.edu.au/hbspapers/70/

Ogle, J. P., & Park, J. (2018). Maternal experiences of parenting girls who are perceived as overweight or at risk for becoming so: Narratives of uncertainty, ambivalence and struggle. *Children & Society, 32*(4), 325–340. https://doi.org/10.1111/chso.12258

Persky, S., McBride, C. M., Faith, M. S., Wagner, L. K., & Ward, D. S. (2015). Mothers' guilt responses to children's obesity risk feedback. *Journal of Health Psychology, 20*(5), 649–658. https://doi.org/10.1177/1359105315576608

Persky, S., Yaremych, H. E., Goldring, M. R., Ferrer, R. A., Rose, M. K., & Hollister, B. M. (2020). Investigating the efficacy of genetic, environmental, and multifactorial risk information when communicating obesity risk to parents of young children. *Annals of Behavioral Medicine, 55*(8), 720–733. https://doi.org/10.1093/abm/kaaa103

Pescud, M., & Pettigrew, S. (2014). "I know it's wrong, but…": A qualitative investigation of low-income parents' feelings of guilt about their child-feeding practices. *Maternal & Child Nutrition, 10*(3), 422–435. https://doi.org/10.1111/j.1740-8709.2012.00425.x

Pocock, M., Trivedi, D., Wills, W., Bunn, F., & Magnusson, J. (2010). Parental perceptions regarding healthy behaviours for preventing overweight and obesity in young children: A systematic review of qualitative studies. *Obesity Reviews, 11*(5), 338–353. https://doi.org/10.1111/j.1467-789X.2009.00648.x

Presaghi, F., Desimoni, M., & Presaghi, M. F. (2019). *random.polychor.pa: A parallel analysis with polychoric correlation matrices. R package version 1.1.4-4*. https://cran.utstat.utoronto.ca/web/packages/random.polychor.pa/random.polychor.pa.pdf

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Ainsworth, A. T., & Haviland, M. G. (2005). Item response theory: Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science, 14*(2), 95–101. https://doi.org/10.1111/j.0963-7214.2005.00342.x

Samejima, F. (2016). Graded response models. In W. J. van der Linden (Ed.), *Handbook of item response theory* (Vol. 1, pp. 123–136). Chapman and Hall/CRC.

Savage, J. S., Fisher, J. O., & Birch, L. L. (2007). Parental influence on eating behavior: conception to adolescence. *The Journal of Law, Medicine & Ethics, 35*(1), 22–34. https://doi.org/10.1111/j.1748-720X.2007.00111.x

Scarnier, M., Schmader, T., & Lickel, B. (2009). Parental shame and guilt: Distinguishing emotional responses to a child's wrongdoings. *Personal Relationships, 16*(2), 205–220. https://doi.org/10.1111/j.1475-6811.2009.01219.x

Schmidt, F. L., & Hunter, J. E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*(2), 199–223. https://doi.org/10.1037/1082-989X.1.2.199

Seagram, S., & Daniluk, J. C. (2002). "It goes with the territory": The meaning and experience of maternal guilt for mothers of preadolescent children. *Women & Therapy, 25*(1), 61–88. https://doi.org/10.1300/J015v25n01_04

Snuggs, S., Houston-Price, C., & Harvey, K. (2019). Development of a parental feeding goal measure: The family mealtime goals questionnaire. *Frontiers in Psychology, 10*, Article 455. https://doi.org/10.3389/fpsyg.2019.00455

Tangney, J. P. (1996). Conceptual and methodological issues in the assessment of shame and guilt. *Behaviour Research and Therapy, 34*(9), 741–754. https://doi.org/10.1016/0005-7967(96)00034-4

Tignor, S. M., & Colvin, C. R. (2017). The interpersonal adaptiveness of dispositional guilt and shame: A meta-analytic investigation. *Journal of Personality, 85*(3), 341–363. https://doi.org/10.1111/jopy.12244

Tignor, S. M., & Colvin, C. R. (2019). The meaning of guilt: Reconciling the past to inform the future. *Journal of Personality and Social Psychology, 116*(6), 989–1010. https://doi.org/10.1037/pspp0000216

Tracy, J. L., & Robins, R. W. (2006). Appraisal antecedents of shame and guilt: Support for a theoretical model. *Personality and Social Psychology Bulletin, 32*(10), 1339–1351. https://doi.org/10.1177/0146167206290212

Vaughn, A. E., Tabak, R. G., Bryant, M. J., & Ward, D. S. (2013). Measuring parent food practices: A systematic review of existing measures and examination of instruments. *International Journal of Behavioral Nutrition and Physical Activity, 10*(1), Article 61. https://doi.org/10.1186/1479-5868-10-61

Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment* (pp. 41–71). Springer.

Wardle, J., Guthrie, C. A., Sanderson, S., & Rapoport, L. (2001). Development of the children's eating behaviour questionnaire. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 42*(7), 963–970. https://doi.org/10.1017/S0021963001007727

Watson, D., & Clark, L. A. (1994). *The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form*. https://doi.org/10.17077/48vt-m4t2

Wilson, M. (2004). *Constructing measures: An item response modeling approach*. Routledge.

Wright, B. D., & Masters, G. N. (1981). *The measurement of knowledge and attitude*. Statistical Laboratory, Department of Education, University of Chicago.

Yaremych, H. (2023). *Development and validation of the Parental Food Choice Guilt Scale* [Materials]. https://osf.io/9mvhp

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Directorate of Human Resources Research and Evaluation, Department of National Defense. https://faculty.educ.ubc.ca/zumbo/DIF/handbook.pdf

History
Received January 17, 2023
Revision received August 6, 2023
Accepted August 12, 2023
Published online November 6, 2023
EJPA Section / Category Sports & Health Psychology

Acknowledgments
The authors thank the expert reviewers, Tom Carpenter, Rebecca Ferrer, Leah Lipsky, Travis Masterson, Daniel Sznycer, and Dianne Ward, for providing thoughtful feedback on our initial pool of scale items. We thank Sun-Joo Cho for providing methodological expertise. We also thank the focus group members who helped to generate items.

Conflicts of Interest
The authors declare that they have no conflicts of interest.

Publication Ethics
Informed consent was obtained from all participants in the study. All waves of data collection were approved by the Institutional Review Board of the National Human Genome Research Institute.

Authorship
Haley Yaremych and Susan Persky conceptualized the scale. S. Persky secured funding. H. Yaremych and S. Persky collected data. H. Yaremych conducted analyses and wrote the manuscript. H. Yaremych and S. Persky edited the manuscript and have approved its final version.

Open Science
We report how we determined our sample size, all data exclusions (if any), all data inclusion/exclusion criteria, whether inclusion/exclusion criteria were established prior to data analysis, all measures in the study, and all analyses including all tested models. If we use inferential tests, we report exact *p* values, effect sizes, and 95% confidence or credible intervals.
Open Data: The information needed to reproduce all reported results is not openly accessible. Data are available on request from the corresponding author.
Open Materials: The information needed to reproduce all reported methodology is not openly accessible. Material is available on request from the corresponding author. All Supplemental Materials referenced in the manuscript are publicly available at https://osf.io/9mvhp/ (Yaremych, 2023).
Preregistration of Studies and Analysis Plans: This study was not preregistered.

Funding
This project was funded by the National Human Genome Research Institute [grant number: Z01-HG200396-05].

Haley E. Yaremych
Department of Psychology & Human Development
Vanderbilt University
PMB 552, 230 Appleton Place
Nashville, TN 37203-5721
USA
haley.e.yaremych@vanderbilt.edu

