Journal of Child and Family Studies
https://doi.org/10.1007/s10826-024-02818-x

ORIGINAL PAPER

The Positive Discipline in Everyday Parenting Questionnaire: Psychometric Characteristics of a Revised Version

Elena Gallitto[1] · Elisa Romano[1]

Accepted: 5 March 2024
© Crown 2024

Abstract

This study describes the development of the Positive Discipline in Everyday Parenting Questionnaire (PDEPQ), which assesses key parental cognitions underlying punitive parental behaviours, namely 1) Approval of Physical Punishment; 2) Approval of Non-Physical Punishment; 3) Subjective Norms; and 4) Parenting Self-Efficacy. In Study 1, two samples (parents, professionals) were recruited to assess the content validity of the four scales, following which a revised version was created. In Study 2, the revised PDEPQ was administered to 400 Canadian parents to examine the factor structure as well its reliability and validity. Results indicated a good factor structure of the scales, but several items were removed to achieve a better fit. Internal consistency of the four scales ranged from acceptable to excellent. Correlations between the scale scores at the two time points were large in magnitude and significant, indicating excellent test-retest reliability. Furthermore, correlations between the PDEPQ scales and measures of similar constructs were significant and in the expected direction, indicating a strong alignment of each scale with its respective construct of interest. In conclusion, findings from these various studies indicate good psychometric properties of the PDEPQ scales, as well as strong evidence of construct validity, internal consistency, and test-retest reliability.

Keywords Physical punishment · Emotional punishment · Parental attributions · Parenting self-efficacy · Psychometric assessment

Highlights

- Psychometric assessment of a novel measurement instrument to assess parents' social cognitions of physical and emotional punitive parenting.
- Validation of a measurement instrument assessing of attitudes toward physical and emotional punishment, social norms, and sense of self-efficacy.
- Use of Items Response Theory analysis in parenting research to assess dimensions of punitive parenting.

Physical punishment is a concerning societal problem that violates children's rights to protection from violence. Solid research shows that physical punishment negatively affects children's adjustment across multiple domains of functioning (e.g., emotional, interpersonal, cognitive) on both a short- and long-term basis (Gershoff, 2002; Gershoff & Grogan-Kaylor, 2016). A meta-analysis of 75 studies that focused exclusively on spanking yielded 79 unique significant effect sizes; of these, 99% indicated associations between spanking and adverse child outcomes, including aggression, antisocial behaviour, externalising and internalising behaviour problems, negative relationships with parents, lower moral internalisation, and slower cognitive development (Gershoff & Grogan-Kaylor, 2016). The effect sizes did not vary based on whether studies were longitudinal, experimental, or cross-sectional in design. Furthermore, the relationships between spanking and child outcomes were similar to those between more severe physical abuse and the same child outcomes. Given the pervasive adverse effects of punitive parental behaviour, examining some critical cognitive components underlying such rearing practices is essential.

✉ Elena Gallitto
egall065@uottawa.ca

[1] School of Psychology, University of Ottawa, Ottawa, ON, Canada

Published online: 29 April 2024

Social Information Processing (SIP; Milner, 2003) theory focuses on some critical cognitive processes underlying parental punitive behaviour. SIP theory suggests that physical force against a child can be explained by a series of parental cognitions, including unrealistic expectations of the child, lack of perceived parental control, and hostile attribution biases (Milner, 2003). These important set of parental cognitions (can also be classified as follows: (1) *evaluative cognitions*, such as attitudes toward child-rearing, which are characterised by pre-existing beliefs regarding the use of specific parenting practices; (2) *analytical cognitions*, including "conflict-promoting attributions" that are centred around attributing responsibility to the child for challenging behaviour; and (3) *self-efficacy cognitions* that involve parents' perceptions of their sense of competence in their parenting role and their disciplinary styles (Bugental & Johnston, 2000).

Given the importance of parental cognitions, it would seem critical to design assessment tools that capture various related aspects, such as attitudes, attributions, and parenting self-efficacy linked to parental punitive behaviour. Therefore, the main objective of this study was to examine the factor structure of the Positive Discipline in Everyday Parenting Questionnaire (PDEPQ), a measurement tool designed to assess critical cognitive components of punitive parental practices. Moreover, we made some revisions to the original version of the instrument and conducted a valid and reliable assessment of the newly designed measure of parental cognitions.

The Positive Discipline in Everyday Parenting Questionnaire

The Positive Discipline in Everyday Parenting questionnaire (PDEPQ) is rooted in SIP theory, which attributes punitive behaviour to punishment-related parental cognitions. Durrant et al. (2014) designed this questionnaire to measure cognitive predictors of physical and non-physical punishment, which are the targets of the Positive Discipline in Everyday Parenting programme (PDEP; for more information, see Durrant, 2007, 2013). The main programme objectives are to reduce parents' support for physical punishment, shift attributions around child behaviours, and improve parenting self-efficacy. The measure was initially developed in 2012 to gather data about PDEP outcomes as the programme evolved and was piloted in many countries, particularly throughout Asia. However, several issues with the initial version of the PDEPQ were identified.

First, there is limited information on the psychometric characteristics of the PDEPQ. Moreover, Durrant et al. (2014) reported that the reliability of the Parenting Self-Efficacy scale is in the low range ($\alpha = 0.49$–0.52). Therefore, a detailed psychometric analysis is required to increase confidence in the primary questionnaire used to evaluate the PDEP programme. This paper presents two studies involving a series of item-level analyses to determine the psychometric properties of the PDEPQ and then apply all the necessary modifications to achieve a more reliable and valid measurement instrument. These two steps resulted in a revised version of the PDEPQ, which can be reliably used in programme evaluation and parenting assessment. The PDEPQ includes the following scales: *Approval of Physical Punishment; Approval of Non-Physical Punishment; Subjective Norms; and Parenting Self-Efficacy*.

Approval of Physical Punishment

The *Approval of Physical Punishment* scale measures the degree to which parents favour using physical punishment (i.e., favourable attitudes) and covers the age range from young children to adolescents. This scale includes items tapping into parental attitudes toward various types of physical punishment, including spanking, slapping, and pinching. Generally, attitudes are "a mental and neural state of readiness, organised through experience, exerting a directive or dynamic influence upon the individual's response to all objects and situations with which it relates" (Allport, 1935, p. 810). In child-rearing practices, attitudes toward physical punishment are critical factors in determining the strategies parents are likely to use (Holden, 2020; Vittrup et al., 2006). For instance, Ateah and Durrant (2005) identified four variables as most predictive of parental physical punishment use, namely (1) approval of physical punishment use, (2) attributions for the child's transgression, (3) perceived seriousness of the child's transgression, and (4) anger in response to the transgression. These four variables (except for parental anger, which is not classifiable as parental cognitions) accounted for 54% of the parents' physical punishment use variance.

Moreover, approval of physical punishment emerged as the strongest predictor, which included other variables such as a mother's childhood history of physical punishment, knowledge of child development, and repertoire of positive disciplinary strategies (Ateah & Durrant, 2005). Perron et al. (2014) studied 2340 Canadian parents of children 2 to 12 years old. They found that after controlling for the child- and family-related factors, a favourable parental attitude toward spanking was the strongest predictor of the use of spanking. In addition, parents who held pro-spanking attitudes were nearly five times more likely to endorse physical punishment.

Interestingly, research on child-rearing attitudes has highlighted a gap in the measurement of parental attitudes toward punishment, partially due to the limited availability

of valid and reliable measures (Holden & Buck, 2002). Current measurement instruments used to assess punitive parenting have limitations in that they focus on spanking and exclude other forms of physical punishment, such as slapping (i.e., hitting any part of a child's body with the open palm), commonly used with older children and adolescents. Moreover, no existing measure is available on attitudes toward non-physical punishment. The PDEPQ is unique because it includes a scale to measure common forms of non-physical punishment, such as verbal threats, ridicule and time-out. Such responses could potentially lead to a child's sense of rejection and isolation and, as such, are considered detrimental in that they might undermine the child's confidence and self-esteem (Gittins & Hunt, 2019).

Approval of Non-Physical Punishment

The *Approval of Non-Physical Punishment* scale measures the degree to which parents favour using non-physical punishment (i.e., time-out, grounding, shaming). This is a construct that has not received much empirical investigation at the assessment and measurement levels. The use of non-physical punishment is a widely debated topic that has fomented interest among researchers in parenting (Holden et al., 2017; Larzelere et al., 2017). The most common type of non-physical punishment is time-out, which is a form of negative punishment that involves removing a child from the situation that caused challenging behaviour and placing them in an environment that is "less" reinforcing for a reasonably deemed amount of time (e.g., usually 2–5 min; Dadds & Tully, 2019). Dadds and Tully's (2019) review highlighted variations in the delivery of time-out, whether inclusionary (e.g., time-out chair) or exclusionary, and how different modalities might impact behavioural outcomes and the child's emotional and social development.

Time-out is presented as an effective strategy to deal with challenging behaviour in several widely used evidence-based programmes, including Triple P (Sanders et al., 2014), Parent–Child Interaction Therapy (Querido et al., 2002), and Incredible Years (Webster-Stratton, 2001). Furthermore, this method has also been explicitly recommended by the American Academy of Pediatrics (2018) as a viable alternative to physical punishment. Despite immediate positive effects of time-out on child compliance that have been proposed as a potentially beneficial aspect of this procedure, other essential features of the context in which these disciplinary measures are implemented (e.g., attachment relationships, communication style, social context) need to be considered to gauge their overall impact on child development (Dadds & Tully, 2019).

Although there are claims about the effectiveness of time-out in addressing children's behavioural challenges (Kaminski & Claussen, 2017; Larzelere et al., 2017), its use as an appropriate (and effective) disciplinary strategy remains debatable. Results from an international study—involving China, India, Kenya, Philippines, and Thailand - with a sample of 292 families indicated that the frequency of use of time-outs was linked to greater internalising symptoms among 8–12 years olds (Gershoff et al., 2010). Moreover, no evidence exists that using time-out is superior to non-behavioural disciplinary techniques in fostering positive developmental outcomes. For instance, a recent longitudinal study reported no differences in the emotional and behavioural functioning of children whose parents used time-outs regularly relative to those who did not engage in this parenting practice. Knight et al. (2020) concluded that time-outs have neutral (non-detrimental) effects on children's psychological well-being. However, the authors did not include positive outcomes in their study, making it difficult to reach a definite conclusion on the beneficial effects of time-outs on children's behaviour. Furthermore, longitudinal studies on the prolonged positive effects of time-out are lacking, pointing to the need for additional research to determine the impact of this parenting technique on children's outcomes. Therefore, it is essential to design measurement instruments that capture other aspects of non-physical punishment, including time-out.

Subjective Norms

The *Subjective Norms* scale of the PDEPQ measures parental attributions of children's behaviours; it includes items for all primary age groups (infancy through to adolescence) and is not constrained to a target child. These are parents' interpretations of children's behaviour as developmentally normative, as opposed to an intentional act on the part of the child (Durrant et al., 2014). For instance, a negative social norm might be the belief that children say "no" to make their parents mad. Such attributions will likely increase the likelihood of punitive parental responses (Holden, 2020). Similarly, parents' negative attributions to children's behaviour have been shown to relate to harsh parenting (Crouch et al., 2017; Sturge-Apple et al., 2020) and an increased likelihood of physical punishment use (Ateah & Durrant, 2005; Rodriguez & Wittig, 2019).

Concerning subjective norms, several parental self-report questionnaires tap into attributions for children's challenging behaviours. Most measures assess the degree to which parents attribute negative intentionality to child misbehaviour (e.g., the child is mad at the parent, wants to embarrass the parent, does not care). However, most existing measures focus on specific child behaviours and exclude underlying assumptions concerning children of various age groups. A novel aspect of the Subjective Norms subscale of the PDEPQ is the focus on parents' negative attributions for typical behaviours of children of different

ages to allow the capture of various ranges of attributions from parents.

Parenting Self-Efficacy Scale

Finally, the *Parenting Self-Efficacy* scale measures parents' sense of self-efficacy and competence in their parenting role. This scale was designed to measure parenting self-efficacy in general and dealing with challenging child behaviours without conflict. Parenting self-efficacy refers to beliefs concerning one's abilities to manage the daily tasks of parenthood and a sense of competence in one's parenting role (Fang et al., 2021). Parenting self-efficacy can also impact how parents respond to conflict with their children. For example, low parenting self-efficacy has been related to frequent use of physical punishment (Khoury-Kassabri et al., 2014; Rominov et al., 2016). On the contrary, high levels of parenting self-efficacy have consistently been associated with competent and positive parenting practices (Hamovitch et al., 2019).

Concerning existing measures of parenting self-efficacy, domain-general self-efficacy focuses on the global assessment of self-efficacy (i.e., overall sense of competence) and is considered independent of domain-specific self-efficacy (i.e., the individual's self-perceived ability in a specific parenting task; Coleman & Karraker, 2000). A systematic review of 76 studies identified 34 self-report measures of parenting self-efficacy (Wittkowski et al., 2017), 22 of which measured domain-specific self-efficacy. These scales were compared on their psychometric qualities (e.g., content and construct validity, reliability) and general characteristics (e.g., ease of scoring, interpretability). Among the scales reviewed, the Self-efficacy for Parenting Tasks Index (SEPTI; Coleman & Karraker, 2000) did not receive a high ranking compared to other domain-specific scales, particularly in terms of content validity (i.e., receiving a score of 1—indicating doubtful methodology) and test-retest reliability (i.e., receiving a score of 0—indicating no information available on the replicability of the score; Wittkowski et al., 2017). Moreover, the Discipline subscale of the SEPTI assesses perceived self-efficacy concerning disciplinary measures but does not reference physical punishment or managing conflict with children. The Parenting Self-Efficacy scale of the PDEPQ is unique in that it includes elements of both global and domain-specific self-efficacy specifically tailored to assess confidence in resolving conflict without using punishment, and it is not constrained to only young children. Therefore, these items are applicable to both young children and adolescents. In addition, there was an effort to avoid language that assumed the child's intent, such as the word 'misbehaviour.' Instead, specific behaviours were described so parents could generate spontaneous responses.

In sum, the four scales that are part of the PDEPQ allow researchers to capture some critical predictors of parental punishment that existing measures lack. In this study, the validity of the PDEPQ was examined in terms of content and construct. Moreover, a detailed psychometric evaluation of the measurement instrument, including a series of item-level analyses and a reliability assessment, was undertaken to ensure the development of psychometrically sound scales designed to measure cognitive dimensions of parenting related to parental punishment.

Study Objectives

Through two studies, we aimed to evaluate the PDEPQ-R (revised version) and gather data on its psychometric properties. Study 1 assessed its content validity via two samples: (1) a group of professionals in parenting research, child development, and psychometric assessment and (2) a group of parents. Following questionnaire modifications based on Study 1 findings, the goal of Study 2 was to gather data from a sample of 400 parents to examine the psychometric properties of the PDEPQ-R using Item Response Theory (IRT) analyses. The sample data were also used to gather evidence to support the construct validity of the four scales and their internal consistency. Finally, a sub-sample ($n = 100$) of parents was randomly selected among the 400 parents who completed the PDEPQ-R to complete the same questionnaire (i.e., no other measure was administered) four weeks following the initial administration to assess its test-retest reliability.

One significant advantage of this study was the combination of Classic Test Theory (CCT) and IRT methodologies to assess the psychometric characteristics of the PDEPQ-R scales. Generally, Cronbach's alpha has been criticised as being population-dependent (i.e., coefficient changes depending on the test-taker sample) and subject to possible inflations with increasing items in a given scale (Reise & Henson, 2003). On the other hand, the IRT approach is not only population-independent (i.e., consistent estimates across different populations), but it also provides a reliability measure for each item (e.g., item information curve) at different levels of the underlying trait so that each item is associated with information parameters (De Ayala, 2013). Thus, the use of IRT, in addition to the traditional reliability assessment (i.e., Cronbach alpha), constitutes an advantage in interpreting the reliability of a measurement instrument in that it provides valuable information at various stages of scale development. In addition, this study used quantitative and qualitative techniques to examine the psychometric characteristics of the PDEPQ-R. These techniques were used sequentially, beginning with the content validity assessment (professional and parent

feedback) and associated modifications, followed by the IRT analyses.

Study 1. Content Validity Analysis of the Revised PDEP Questionnaire

A content-based validity assessment of the PDEPQ-R evaluated whether the items composing the four scales were consistent with the constructs that they were intended to measure.

Methods

Participants and Procedures

The first sample comprised eight professionals from Canada and the U.S. whom we identified as having expertise in child development, parenting research, programme evaluation, and psychometric assessment. This list initially included 12 individuals contacted by email to participate as content experts for the PDEPQ-R. The sample of eight professionals was predominately female (87.5%) and White (87.5%), with a mean age of 44.9 years ($SD = 12.2$).

The second sample was composed of 10 parents recruited through two local community centres offering parenting services. Participants were included in the study if they were adults who reported being comfortable with the English language. Two focus groups ($n = 5$ each) were held, and this sample size of 10 is consistent with similar research employing focus groups as part of a content validity assessment (Barnes & Adamson-Macedo, 2007; Marsteller et al., 2017). Each parent focus group lasted about two hours. In each group, an overview of the study was presented, and then parents completed a consent form, a sociodemographic questionnaire, and a content validity assessment form. Subsequently, the parents took part in the focus group discussion. The focus group was not audiotaped to increase parents' comfort level; however, field notes were taken. For both objectives (i.e., expert panel and focus group), approval was obtained from our university's Office of Research Ethics and Integrity before study commencement. Consent for the expert panels was sought when the online questionnaire was administered, whereas physical copies of consent were provided to parents participating in the focus group.

The sample was composed exclusively of mothers whose average age was 35.2 ($SD = 4.6$) years. Most participants (75%) were married. One-third of the participants indicated that their ethnic/racial background was Arab/Middle Eastern (33.3%), followed by White (16.7%), Latin American (16.7%;), South Asian (16.7%), West Asian (8.3%), and Black (8.3%). Most mothers completed post-secondary education (66.7%) and were not employed outside the home (66.7%). Household income ranged from $20,000-$49,999 (40%) to $50,000-$89,999 (20%) and $90,000–109,999 (40%) in Canadian dollars.

Measures

PDEP questionnaire (PDEPQ)

The first column of Table 1 shows the version of the PDEPQ used as the starting point for the content validity evaluation. It is composed of four scales. *Approval of Physical Punishment* consists of 13 items on parental attitudes toward physical punishment. Each item is rated along a 6-point Likert scale from 1 (*strongly disagree*) to 6 (*strongly agree*), with higher scores indicating greater approval of physical punishment. *Approval of Non-Physical Punishment* consists of 10 items on parental attitudes toward non-physical punishment. Each item is rated along a 6-point Likert scale from 1 (*strongly disagree*) to 6 (*strongly agree*), with higher scores indicating greater approval of non-physical physical punishment. The *Subjective Norms scale* has ten items to assess parents' negative attributions of typical child behaviours. Participants are asked to indicate their agreement with statements attributing typical parent-child conflict to intent on the part of the child. Parents respond to each item on a 6-point Likert scale from 1 (*strongly disagree*) to 6 (*strongly agree*), with higher scores indicating a stronger belief that typical child behaviour is attributable to intentional misbehaviour. *Parenting Self-Efficacy* consists of 9 items that measure the degree to which parents feel competent and confident in their ability to manage challenging parenting situations. Each item is rated along a 5-point Likert scale from 1 (*never*) to 5 (*always*), with higher scores indicating greater parenting self-efficacy.

For the content validity assessment, the items of each of the four scales were rated on their representativeness (i.e., how reflective the item is of the underlying construct) using a 3-point scale from 1 (*slightly representative)* to 3 (*highly representative*) and their clarity (i.e., how clear the item is) using a 3-point scale ranging from 1 (*low clarity*) to 3 (*high clarity*). Participants were also asked to provide any suggestions they might have on how to improve the items.

Data Analysis

Descriptive analyses were employed to examine mean representativeness and mean clarity for each item using SPSS version 26. We also considered participants' comments by way of focus group field notes that were examined using a note-based analysis, which is appropriate when the purpose of a study is narrowly defined (e.g., pilot-testing

Table 1 Changes to PDEP Items Based on the Content Experts' and Parents' Feedback

Approval of Physical Punishment	Initial Item	Modifications
APP1	Sometimes a spank or slap is the best way to get a young child to listen.	
APP2	Spanking or slapping is fine as long as the parent is not angry.	
APP3	Parents should have the rights to decide whether to spank their children.	
APP4	If parents don't spank or slap their children, they will be spoiled.	
APP5	It's ok to spank a 5-year old's bottom if she does something dangerous.	
APP6	It's ok to slap a teenager's face if she uses bad words in front of her parents.	It's ok to slap a teenager's face if she **insults her parent**.
APP7	Young children who aren't spanked will get into trouble when they are teenagers.	
APP8	Spanking teaches children right from wrong.	
APP9	If a 4-year-old won't leave a store, his parent should pull his arm and drag him out.	If a 4-year-old won't leave a store, his parent should **pull him** out by his arm.
APP10	It's ok to shake a 3-year-old to make her stop screaming.	
APP11	If a 2-year-old bites another child, it's ok to bite him to teach him how it feels.	
APP12	Children who are spanked learn to respect their parents	
APP13	It's ok to slap a 2-year-old's hand for touching dangerous objects.	It's ok to slap a 2-year-old's hand for touching an **object he was told not to touch**.
Approval of Non-Physical Punishment		
ANP1	If a 14-year old is failing school, his parent should make him do extra chores.	If a 14-year-old is failing in school, his parent should make him do extra chores **to teach him to work harder**.
ANP2	If a 12-year-old gets into trouble at school, her parent should punish her before she has a chance to give excuses.	If a 12-year-old gets into trouble at school, her parent should **scold** her before she has a chance to give excuses.
ANP3	If a 16-year-old wears a hairstyle that his parent disapproves of, he should have to stay inside until he changes it.	If a 16-year-old wears **clothing** that his parent disapproves of, he should have to stay inside until he changes it.
ANP4	If children break the rules, their parents should take away privileges.	If a 10-year-old breaks the rules, her parents should take away favourite activities**, like watching TV or seeing friends**.
ANP5	If a 3-year-old won't eat, his parent should make him stay at the table until he finishes his food.	
ANP6	If two sisters won't stop fighting, the parent should threaten to take away their phones.	If **two teenage siblings** won't stop fighting, their parent should threaten to take something away, such as their phones.
ANP7	A good way to teach a 13-year-old to dress properly is to embarrass her in front of her friends.	
ANP8	If a 5-year-old won't get dressed, his parent should threaten to take him out in his underwear.	If a 6-year-old won't get dressed for school, his parent should **force the shirt over his head**.
ANP9	A 10-year-old who hurts another child should be forced to apologise.	A **7-year**-old who hurts another child should be forced to apologise.
ANP10	If a 3-year-old won't share his toys, he should be made to sit on a chair for 3 min.	
Subjective Norms		
SN1	Two-year-olds who say "no!" are being defiant.	
SN2	Usually, 3-year-olds have tantrums because they are spoiled.	
SN3	Four-year-olds who interrupt adults are rude.	Four-year-olds who interrupt adults **should know better**.
SN4	A 13-year-old who does not want to be seen with his mother should be ashamed of himself.	

Table 1 (continued)

Approval of Physical Punishment	Initial Item	Modifications
SN5	If an 8-year-old uses bad words in front of his parents, this is a sign of disrespect.	
SN6	A 16-year-old who questions her parents' religious beliefs is disrespectful.	
SN7	A 5-year-old child who won't go to sleep is being stubborn.	
SN8	A 6-year-old child who won't show affection to his grandmother is rude.	
SN9	Teenagers who live at home should always follow their parents' rules, even if they don't agree	Teenagers **should follow their parents'** rules without question, even if they don't agree.
SN10	Boys should learn not to cry by the time they are 12 years old.	**Deleted**
Parenting Self-Efficacy		
SE1	I can solve challenging situations without punishing this child.	
SE2	I believe I have the skills to be a good parent to this child.	
SE3	I feel like I just don't know what to do with this child.	
SE4	I can control my anger in challenging situations with this child.	
SE5	I feel badly about how I handle challenging situations with this child.	
SE6	I feel hopeful that I can improve my relationship with this child.	I feel **confident** that I can improve my relationship with this child.
SE7	In challenging situations, I end up shouting at this child.	In challenging situations, **I am able to stop myself from shouting at this child**.
SE8	I believe that I will be able to handle future challenges as this child gets older.	
SE9	I feel hopeless about my relationship with this child	

APP Approval of Physical Punishment, *ANP* Approval of Non-Physical Punishment, *SN* Subjective Norms, *SE* self-efficacy

items; Krueger, 2014). The first author read all the notes to understand the main ideas communicated by the participants. Next, codes were generated to capture themes, and then similar participant comments were assigned to the same codes.

Results

Experts' mean ratings of the representativeness of the Approval of Physical Punishment items were high ($M = 2.8$; $SD = 0.3$), whereas parents rated the items as moderately representative ($M = 2.5$; $SD = 0.5$). The mean clarity rating for professionals and parents was 2.7 ($SD = 0.5$), indicating high clarity for the items on this scale. The language in item APP6 was flagged as problematic because the term "bad words" was too specific. The term "drag him out" in item APP9 was considered inappropriate by several professionals and parents. Moreover, professionals and parents highlighted the term "dangerous" in item APP13 as lacking contextual information. See Table 1 for additional details on the flagged problematic items during this phase.

The mean representativeness of the Approval of Non-Physical Punishment items was high for professionals ($M = 2.7$; $SD = 0.3$) and parents ($M = 2.6$; $SD = 0.4$). Mean clarity ratings were also high among professionals ($M = 2.6$; $SD = 0.5$) and parents ($M = 2.7$; $SD = 0.4$). However, parents reported several concerns with item wording. They indicated that the term "extra chores" in item ANP1 was too general. Both professionals and parents also indicated that the term "punish" in item ANP2 was too general and could be misinterpreted as physical punishment. The term "hairstyle" in ANP3 was also noted as odd. Both professionals and parents indicated that item ANP4 was too general and could include children's favourite activities (e.g., electronics, sleepovers) instead. Several parents noted Item ANP6, as sibling rivalry is not exclusive to one sex. Most parents (and several professionals) agreed that item ANP8 was too harsh and implied humiliation and should be changed to reflect more inductive forms of punishment.

Parents indicated that this item could be reworded to match the definition of non-physical punishment (i.e., taking away privileges).

Subjective Norms items' mean representativeness was high for professionals ($M = 2.7$; $SD = 0.3$) and parents ($M = 2.6$; $SD = 0.6$). Mean clarity ratings were also high among professionals ($M = 2.7$; $SD = 0.5$) and parents ($M = 2.8$; $SD = 0.4$). The term "are rude" in SN3 was substituted with an alternative phrase. One parent expressed difficulties understanding some terminology (e.g., defiant in SN1, tantrum in SN2). Most parents thought Item SN10 reflected a gender stereotype and suggested removing it.

The two samples rated Parenting Self-Efficacy items as moderately representative ($M = 2.5$; $SD = 0.6$). Parents' mean representativeness was high at 2.8 ($SD = 0.5$). Mean clarity was moderate for both professionals ($M = 2.5$; $SD = 0.2$) and parents ($M = 2.9$; $SD = 0.4$). There were concerns by professionals that items SE6 and SE9 were not representative of the self-efficacy construct.

Conclusion

Overall, professionals and parents provided evidence that the items adequately represented and clearly stated the four PDEPQ scales. However, the two participant groups also made several suggestions for rewording. We reviewed the questionnaire items considering the feedback and made several changes, especially when the feedback was similar across both participant groups. The revised version of the PDEPQ is outlined in the second column of Table 1. It is also important to note that each PDEPQ scale was retained as a unidimensional measure of separate constructs. Therefore, scale scores are not meant to be summed or combined into a total score.

Study 2. Examining the Psychometric Properties of the Revised PDEP Questionnaire

This second study had two objectives. The first objective was to assess the revised PDEPQ (PDEPQ-R) psychometric properties using Item Response Theory (IRT) methodology. The second objective was to gather information on the convergent validity and reliability of the PDEPQ-R after incorporating changes from the IRT analyses.

Methods

Participants and Procedures

Data collection was conducted online. Parents were recruited across several community organisations in Canada through study links posted on agency websites and social media platforms. Inclusion criteria were: (1) having at least one child between birth and 18 years; (2) being at least 18 years of age; (3) having access to a computer; and 4) being able to read English. Parents who agreed to participate in the study were invited to enter a draw for one of four $40 gift cards. Online recruitment for the first part of the study ended after achieving the minimum required sample size necessary to carry out our analyses with sufficient statistical power ($n = 400$; Hulin et al., 1983). A subsample of parents ($n = 100$) was randomly selected and then contacted via email after four weeks of the initial assessment for the test–retest reliability check.

The sample of 400 parents was predominantly female (93.8%) and White (92.8%). The mean age was 35.2 years ($SD = 4.6$), and the majority (68.5%) were married. The number of children in the home ranged from 1–5, with a mean of 2.0 ($SD = 0.9$) and a mean age of 7.1 years ($SD = 4.9$). Most parents had obtained a post-secondary degree (81.6%), and (62.8%) were employed outside the home. Household income was distributed as follows, in Canadian dollars: Less than $20,000 (10.3%); $30,000-$49,999 (14.5%); $50,000–$89,999 (33.6%); $90,000–119,999 (21.3%); $120,000–150,000 (10.3%); and over $150,000 (10%).

The subsample of 100 parents who participated in the test-retest reliability portion of the study was also predominantly female (94%) and White (92.8%). Their mean age was 36.3 years ($SD = 7.3$), and the majority (67%) were married. The number of children in the home ranged from 1–5 ($M = 1.9$, $SD = 0.8$), and the mean child age was 8.1 years ($SD = 6.7$). Most parents had obtained a post-secondary degree (86.0%) and were employed outside the home (63%). Household income was distributed as follows, in Canadian dollars: Less than $29,999 (13%), $20,000-$49,999 (17%), $50,000-$89,999 (34%), and $90,000–119,999 (14%), $120,000-$150,000 (6%) and over $150,000 (16%).

Measures

Revised PDEP Questionnaire

The PDEPQ-R has four scales: Approval of Physical Punishment, Approval of Non-Physical Punishment, Subjective Norms, and Parenting Self-Efficacy. A total of 14 items were modified, and one was deleted (see Table 1).

Attitudes Toward Spanking Scale (ATS; Holden et al., 1995)

This scale consists of 10 statements (e.g., Sometimes a spank is the best way to get my child to listen) rated on a 7-point scale from 1 (*strongly disagree*) to 7 (*strongly*

agree). The ATS measures parents' attitudes toward physical punishment, with a higher score indicating a more positive attitude. The measure has demonstrated good internal consistency in a sample of college-educated mothers ($\alpha = 0.81$–0.91; Holden et al., 1995) and the current sample of 400 parents ($\alpha = 0.87$).

Parent Cognition Scale (PCS; Snarr et al., 2009)

The PCS consists of 30 items to assess how parents endorse child-responsible or parent-causal attributions for challenging child behaviour. Respondents are asked to think about their child's challenging behaviour over the past two months and rate possible causes on a 6-point scale from 1 (*always true)* to 6 (*never true)*. Items are reverse-scored so that higher scores indicate greater endorsement of the attributions (either child-responsible or parent-causal). There are nine child-responsible attributions (e.g., My child tries to get my goat or push my buttons), seven parent-causal attributions (e.g., I'm not patient), and 14 distractor items that are not used in the scoring (e.g., My child is in a stage). For this study, the Child-Responsible Attribution subscale was used to assess convergent validity because of its conceptual similarity with the Subjective Norms scale of the PDEPQ-R. The Child-Responsible subscale has been found to possess good internal consistency in a community sample of couples ($\alpha = 0.81$; Snarr et al., 2009). In addition, the Child-Responsible subscale had excellent internal consistency in the current sample of 400 parents ($\alpha = 0.90$).

Self-Efficacy for Parenting Tasks Scale—Discipline subscale (SEPTI; Coleman & Karraker, 2000)

The Discipline subscale of the SEPTI includes eight items (e.g., I have more trouble with discipline than any other aspect of parenting) measured along a 6-point scale ranging from 1 (*strongly disagree*) to 6 (*strongly agree*). It assesses perceived self-efficacy concerning disciplinary measures and has been found to possess high internal consistency in a sample of mothers ($\alpha = 0.86$; Coleman & Karraker, 2000). In addition, we found good internal consistency for this subscale in the current sample of 400 parents ($\alpha = 0.86$).

Data Analyses

IRT Analyses

Before the IRT analyses, the PDEPQ-R items were examined for the accuracy of data entry, missing values, and fit between their distributions and the assumptions of multivariate analysis. Samejina's (1969) Graded Response Model (GRM) was used as the primary response function model for the IRT analyses. GRM is appropriate for polytomous data, such as Likert-type items with an underlying response continuum. IRTPRO 4.20 (Cai et al., 2011) was used to conduct these analyses with Maximum Likelihood (ML) as the default for the estimation method (Cai et al., 2011) for each of the four PDEPQ-R scales, unidimensionality and local dependence were examined. Unidimensionality was assessed by comparing the ratio of the first with the second eigenvalue in principal component analysis, with values greater than 2 indicating unidimensionality (Hambleton et al. 1991; Lord, 1980). Moreover, the variance explained by the first factor should be at least 20% of the total variance (Reckase, 1979). This test was conducted prior to the IRT analyses to ensure unidimensionality holds. Local dependence (i.e., dependence among items) was assessed using the LD- χ2statistic, with values exceeding 10 indicating a violation of local dependence (Chen & Thiessen, 1997).

Following an assessment of model dimensionality, the model-level fit was assessed through the goodness-of-fit statistics that provide information concerning discrepancies in responses from the respective parameters. Hence, the Root Mean Square Error of Approximation (RMSEA; Steiger & Lind, 1980) was used to assess the model's fit with values lower than 0.06 for a close fit of the statistical model (Hu & Bentler, 1999). The Akaike information criterion (AIC; Akaike, 1974), Bayesian information criterion (BIC; Schwarz, 1978), and M_2 (Maydeu-Olivares & Joe 2005; Maydeu-Olivares & Joe, 2006) were used to compare competing non-nested models. Lower values of these indices indicate a better fit.

Several parameters and statistics were also examined at the item level. Item retention was based on concurrent consideration of all the available indexes. Discrimination parameters are equivalent to factor loadings and represent the strength of the relationship between the item and the latent trait. Discrimination is as follows: 0.01–0.24 is 'none'; 0.25–0.64 is 'low'; 0.65–1.34 is 'moderate'; 1.35–1.69 is 'high'; and >1.7 is 'perfect' (Baker, 2001). Concerning the Item Information Curves (IIC), a visual inspection of the item information functions is performed. Typically, narrower and higher curves imply higher accuracy, whereas shorter and wider curves imply lower accuracy. Generally, very low-height IICs may indicate a poor fit of the item and, as such, could be used to make judgments regarding removing any specific item (De Ayala, 2013).

Reliability and Construct Validity Analyses

Internal consistency was assessed by way of Cronbach's alpha coefficients. The following cut-offs were used: >0.9 is excellent, >0.8 is good, >0.7 is acceptable, >0.6 is questionable, >0.5 is poor, and <0.5 is unacceptable (George &

Mallery, 2003). Test-retest reliability was assessed by determining the correlation between scores on the PDEPQ-R measured twice over four weeks. For the test-retest section of this study, only the PDEPQ-R questionnaire was administered. Pearson correlation coefficients were also used to assess convergent validity. Correlations were examined between both the Approval of Physical Punishment and Approval of Non-Physical Punishment scales from the PDEPQ-R and the Attitudes Toward Spanking scale (ATS; Holden et al., 1995); the PDEPQ-R Subjective Norms scale and the Parental Cognition scale (PCS; Snarr et al., 2009); and the PDEPQ-R Parenting Self-Efficacy scale and the Discipline subscale of the Self-Efficacy for Parenting Tasks Scale (Coleman & Karraker, 2000). The following cut-off values were used: $r = 0$–0.19 is a very weak correlation; $r = 0.20$–0.39 is a weak correlation; $r = 0.40$–0.59 is a moderate correlation; $r = 0.60$–0.79 is a strong correlation; and $r = 0.80$–1 is a very strong correlation (Mueller, 1997).

Results

Data Screening

Data were initially screened for coding errors by checking the range of values and data entry errors. There were no missing data or significant departures from multivariate normality, except for two items from the Approval of Physical Punishment scale, APP6 (Skew = 3.6; Kurtosis = 14.1) and APP10 (Skew = 6.3; Kurtosis = 50.7). These two items had minimal variability across response categories (87–94% of parents strongly disagreed with these items). For the Approval of Non-Physical Punishment scale, ANP7 (Skew = 3.4; Kurtosis = 14) had low variability across response categories (85% of parents strongly disagreed with this item). Attempts were made to render these items less skewed by using simple transformations (i.e., collapsing some low-frequency categories) and more complex methods (i.e., transformations). However, the programme did not provide a viable solution whenever these items were entered into the IRT model (i.e., no appropriate statistics were produced). Therefore, these three items were removed from further analyses.

Testing the Assumptions of Unidimensionality and Local Dependence

The results of the principal component analysis supported a primary dimension for the four PDEPQ scales. The first three eigenvalues for each scale were as follows: 6.05, 0.98, and 0.83 for Approval of Physical Punishment; 2.81, 1.32, and 0.91 for Approval of Non-Physical Punishment; 2.71, 1.28, and 0.91 for Subjective Norms; and 4.37, 0.69, and 0.64 for Parenting Self-Efficacy. Lord's condition for essential unidimensionality was met by all scales, as demonstrated by the greater-than-two ratio of the first to second eigenvalue (Hambleton et al., 1991; Lord, 1980; Reckase, 1979). Specifically, the first to second eigenvalue ratio was 6.17 for Approval of Physical Punishment, 2.12 for Approval of Non-Physical Punishment, 3.42 for Subjective Norms, and 6.33 for Parenting Self-Efficacy. Moreover, the first factor of each scale accounted for more than 20% of the variance, supporting the unidimensional nature of all four scales (Reckase, 1979). The variance accounted for by the primary dimension of the four scales was 55.03% for Approval of Physical Punishment, 28.03% for Approval of Non-Physical Punishment, 30.05% for Subjective Norms, and 48.54% for Parenting Self-Efficacy (Fig. 1).

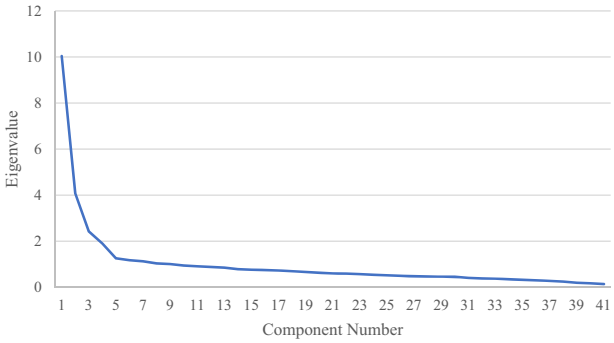


Fig. 1 Principal Component Analysis Scree Plot

Table 2 Model Fit Indices for the PDEP Scales after Revisions

	AIC	BIC	RMSEA	M^2
APP (11 items)	9,254.22	9,513.50	0.060	2911.48
APP (10 items)	8,063.09	8,298.44	0.060	2544.03
APP (9 items)	6,974.06	7,185.43	0.060	1882.05
ANP (9 items)	10,547.11	10,762.52	0.050	1813.16
ANP (7 items)	8,136.88	8,304.42	0.040	811.46
SN (9 items)	9,513.25	9,728.65	0.050	1244.53
SE (9 items)	6,575.93	6,751.44	0.050	534.73
SE (8 items)	5,902.10	6,058.01	0.040	298.07

AIC Akaike Information Criterion, *BIC* Bayesian Information Criterion, *RMSEA* Root Mean Square Error of Approximation, M^2 Maydeu-Olivares and Joe's statistic

IRT Results for the Approval of Physical Punishment Scale

The revised version of the Approval of Physical Punishment scale (11 items) showed a good model fit, as indicated by the lower AIC, BIC, and RMSEA values (See Table 2). As shown in Table 3, most discrimination values for individual items were above 1.70, indicating a perfect fit of the items

Table 3 Discrimination and Threshold Parameters for the PDEP Scales Items

Item Code	Wording	a	b1	b2	b3	b4	b5	M	r
APP1	Sometimes a spank or a slap is the best way to get a young child to listen.	4.40	0.88	2.88	4.28	4.34	7.82	2.23	0.85
APP2	Spanking or slapping is fine as long as the parent is not angry.	5.33	1.05	3.35	5.18	5.55	7.91	2.25	0.87
APP3[a]	**Parents should have the right to decide whether to spank or slap their children**.	2.29	−0.69	−0.14	0.30	0.59	1.36	3.23	0.54
APP4	If parents don't spank or slap their children, they will be spoiled.	3.66	2.54	3.47	4.63	4.84	5.67	1.47	0.77
APP5	It's ok to spank a 5-year-old's bottom if she does something dangerous.	2.53	−0.14	1.05	1.75	2.24	3.91	2.64	0.67
APP7	Young children who aren't spanked will get into trouble when they are teenagers.	3.26	2.63	4.35	5.10	5.15	7.63	**1.29**	0.86
APP8	Spanking teaches children right from wrong.	4.97	1.06	3.36	4.89	5.07	8.33	1.54	0.64
APP9	**If a 4-year-old won't leave a store, his parent should pull him out by his arm.**	1.04	−1.25	2.45	2.18	3.75	4.91	2.22	0.85
APP11	If a 2-year-old bites another child, it's ok to bite them to teach them how it feels.	2.53	−0.14	1.05	1.75	2.24	3.91	2.94	0.75
APP12	Children who are spanked learn to respect their parents.	4.83	1.03	3.08	4.88	5.15	7.63	**1.07**	0.64
APP13	It's ok to slap a 2-year-old's hand for touching an object he was told not to touch.	2.28	−0.94	0.54	1.57	1.78	3.93	2.64	0.76
ANP1	If a 14-year-old is failing in school, his parent should make him do extra chores to teach him to work harder.	1.00	−0.67	0.71	1.75	1.88	3.61	2.17	0.49
ANP2	If a 12-year-old gets into trouble at school, her parent should scold her before she has a chance to give excuses.	1.19	−0.17	1.38	3.18	3.24	4.72	2.81	0.58
ANP3	If a 16-year-old wears clothing that his parent disapproves of, he should have to stay inside until he changes it.	1.11	−1.79	−0.54	0.46	0.70	2.32	2.41	0.62
ANP4	**If a 10-year-old breaks the rules, her parents should take away favourite activities, like watching TV or seeing friends.**	0.89	−4.01	−2.79	−1.80	−0.54	0.71	1.93	0.32
ANP5	If a 3-year-old won't eat, his parent should make him stay at the table until he finishes his food.	1.46	−0.33	1.10	2.41	2.43	4.15	3.36	0.55
ANP6	If two teenage siblings won't stop fighting, their parent should threaten to take something away, such as their phones.	0.82	−3.38	−2.14	−1.27	−0.18	1.14	4.73	0.53
ANP8	If a 6-year-old won't get dressed for school, his parent should force the shirt over his head.	1.15	−2.03	−0.57	0.40	0.51	2.55	2.21	0.61
ANP9	**A 7-year-old who hurts another child should be forced to apologise**.	0.66	−3.21	−2.16	−1.50	−0.16	0.98	4.43	0.48
ANP10	If a 3-year-old won't share his toys, he should be made to sit on a chair for 3 min.	1.65	−0.54	0.97	2.13	2.23	3.79	**1.22**	0.58
SN1	Two-year-olds who say "No!" are being defiant.	1.78	−0.83	1.33	2.59	2.64	4.32	3.42	0.65
SN2	Usually, 3-year-olds have tantrums because they are spoiled.	1.77	−0.22	1.95	3.21	3.35	5.61	4.49	0.61
SN3	Four-year-olds who interrupt adults should know better.	1.38	−1.54	0.52	1.82	1.87	3.87	2.41	0.63
SN4	A 13-year-old who does not want to be seen with his mother should be ashamed of himself.	1.30	−0.87	0.65	2.09	2.27	4.36	2.29	0.61
SN5	If an 8-year-old uses bad words in front of his parents, this is a sign of disrespect.	1.11	−1.69	−0.25	0.91	1.15	2.78	1.96	0.59
SN6	A 16-year-old who questions her parents' religious beliefs is disrespectful.	1.47	0.59	2.04	3.28	3.32	5.30	1.83	0.57
SN7	A 5-year-old child who won't go to sleep is being stubborn.	1.37	−1.76	0.13	1.46	1.52	3.73	2.61	0.62
SN8	A 6-year-old child who won't show affection to his grandmother is rude.	2.05	0.38	2.28	3.63	3.76	7.03	1.74	0.66
SN10	Teenagers should follow their parents' rules without question, even if they don't agree.	1.00	−1.58	−0.20	1.19	1.33	2.81	2.39	0.56
SE1	I can solve challenging situations without punishing this child.	1.88	−8.02	−4.47	−1.35	2.60	–	2.83	0.69
SE2	I believe I have the skills to be a good parent to this child.	1.62	−6.81	−5.68	−2.65	−0.46	–	2.98	0.68
SE3	I feel like I just don't know what to do with this child.	2.31	−6.65	−3.42	−0.24	2.56	–	3.09	0.74
SE4	I can control my anger in challenging situations with this child.	1.88	−7.42	−5.28	−2.47	1.82	–	4.04	0.53
SE5	I feel badly about how I handle challenging situations with this child.	2.21	−4.46	−0.29	3.63	5.94	–	3.81	0.64
SE6	**I feel confident that I can improve my relationship with this child**.	**1.17**	**−2.90**	**−1.51**	**0.47**	**1.38**	–	2.60	0.76
SE7	In challenging situations, I am able to stop myself from shouting at this child.	1.73	−2.12	0.77	4.27	7.28	–	4.48	0.69
SE8	I believe that I will be able to handle future challenges as this child gets older.	2.12	−6.92	−5.23	−2.03	1.64	–	4.47	0.76
SE9	I feel hopeless about my relationship with this child.	2.35	1.28	3.94	6.16	9.22	–	2.24	0.72

[a]Item in bold was deemed problematic and therefore deleted

a = item slope or discrimination parameter; b = between category threshold (difficulty parameter), b1–b5 = item thresholds – each representing categories in a rating scale ranging from 1- to 6 (1-g thresholds are required; where g is the number of categories in a rating scale); M = item mean; r = item-total correlation

APP Approval of Physical Punishment; *ANP* Approval of Non-Physical Punishment; *SN* Subjective Norms; *SE* Parenting Self-Efficacy

(Baker, 2001). The only exception was item APP9, which presented a moderate discrimination value (a = 1.04). The highest discrimination value was recorded for item APP2 (b = 5.46), indicating a strong relation of this item with the latent trait. The spread of category thresholds (−0.14 to 7.91) across different items seemed reasonably broad, implying a broad amplitude of coverage in measuring the latent trait. The IICs of APP2 and APP8 were noticeably

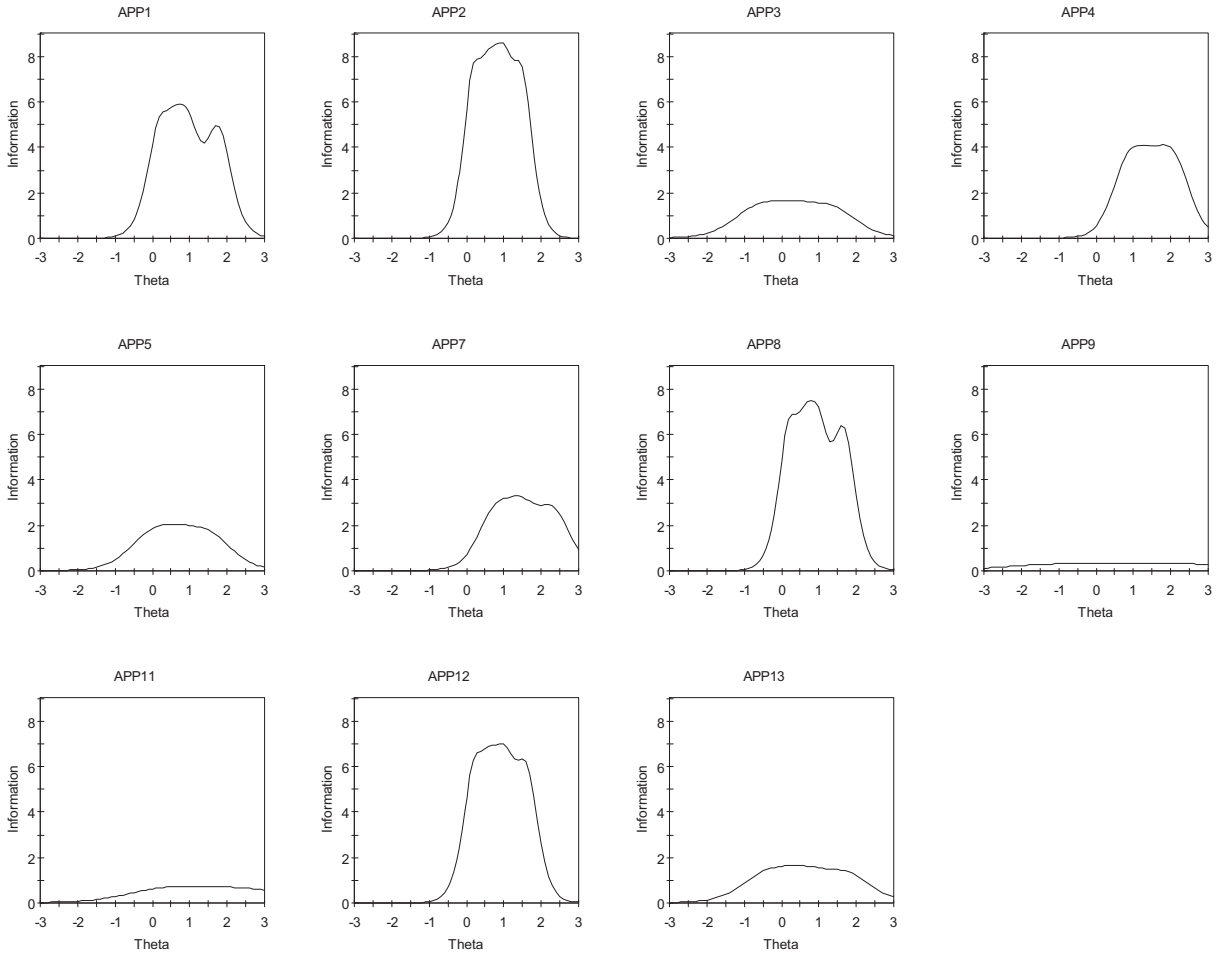


Fig. 2 Information Functions for Approval of Physical Punishment Items

higher than all the other items, suggesting high reliability for these two items (see Fig. 2). The lowest IIC was detected for item APP9, indicating high levels of measurement error for this item compared to other scale items. This item was deleted, and the model was rerun. The model fit improved as indicated by lower AIC, BIC and M^2 values (see Table 2). Most of the LD-χ2 values for the items were acceptable (<10), except for item APP3, which shared unwanted variance with most of the scale items (range LD-χ2 from 12 to 29.3). Hence, item APP3 was removed, and the IRT model was rerun. The reduced model with nine items showed a slightly better fit than the model with 11 items and ten items (APP9 deleted), as indicated by lower goodness-of-fit statistics such as the AIC, BIC, M_2, and RMSEA (see Table 2).

IRT Results for the Approval of Non-Physical Punishment Scale

For the revised version of the Approval of Non-Physical Punishment scale (10 items), there was a good model fit, as demonstrated by low AIC, BIC, and M^2 values and RMSEA values below the recommended cut-off (<0.06; See Table 2). As shown in Table 3, most discrimination values (i.e., *a* slope parameters) for individual items were moderate and ranged between 0.66 and 1.65 (Baker, 2001), implying moderate similarity among items in measuring the latent trait. Examining the statistics for each item revealed an ascending trend from response categories *strongly disagree* (threshold b1) to agree (threshold b5) strongly and covered a wide range of the latent trait (range from −0.33 to 4.72). The highest discrimination value was obtained for ANP10, indicating a strong relationship between this item and the latent construct. Moreover, the IIC of ANP10 was noticeably higher than all the other items, indicating the high reliability of this item.

Conversely, ANP9 received the lowest discrimination value, indicating a weak relationship with the latent trait compared to other items on the same scale. The lowest IIC was also detected for ANP9, which signifies a high level of measurement error for this item (see Fig. 3). Most LD-χ2 values for the items were acceptable (<10),

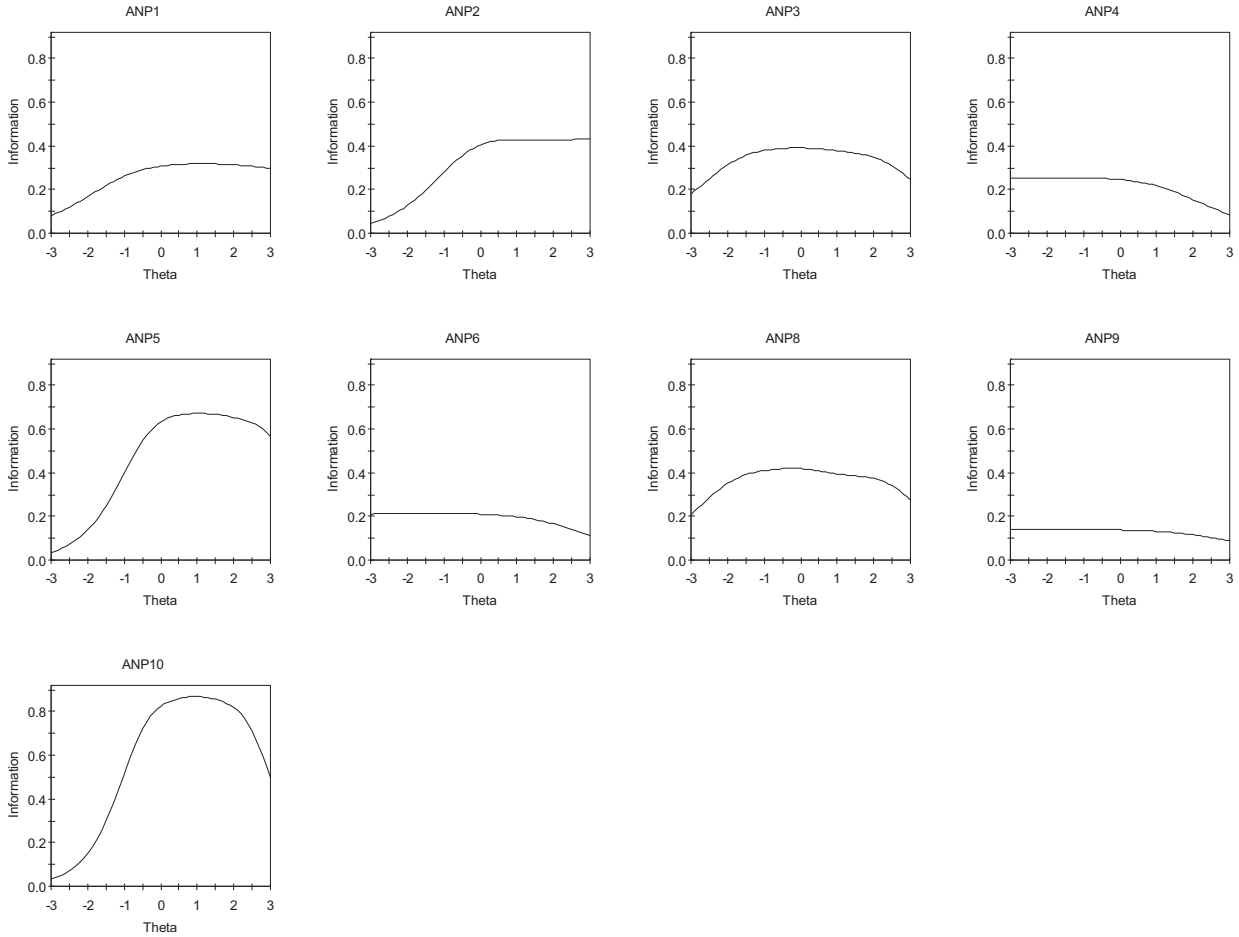


Fig. 3 Information Functions for Approval of Non-Physical Punishment Items

except for ANP4, which shared unwanted variance with ANP3, ANP6, and ANP9 as indicated by LD-χ2 values above the cut-off (LD-χ2 range 24.5–11.00). To address the problem of local dependence between the items, ANP4 was eliminated from the model and rerun. Removing ANP4 and eliminating ANP9 (which had low discrimination and information) improved the model's fit, as demonstrated by lower goodness-of-fit statistics that included the AIC, BIC, M_2, and RMSEA (see Table 2).

IRT Results for Subjective Norms Scale

The model fit statistics of the Subjective Norms scale (9 items) indicated a good fit, as demonstrated by lower AIC and BIC values and an RMSEA value below the recommended cut-off (<0.06; See Table 2). Discrimination values (i.e., the slope for individual items) ranged from 1.00 to 2.05, indicating a moderate to strong relationship to the latent trait (see Table 3). SN8 demonstrated a higher slope parameter, suggesting that this item has a stronger relationship with the latent trait and provides more information about different latent trait levels. All the LD-χ2 values for the items were acceptable (<10), excluding local dependence. The range of thresholds along the latent trait was broad for most items (overall range from −0.22 to 7.03) and in the correct order, implying that most items could measure a reasonable range of the latent trait. The IIC for SN8 was slightly higher than all the other items, indicating the high reliability of this item (see Fig. 4). However, all IIC values were approximately in the same range with no exceptionally shallow curves, suggesting moderate reliability for all the scale items.

IRT Results for the Parenting Self-Efficacy Scale

For the revised version of the Parenting Self-Efficacy scale (9 items), there was a good model fit, as demonstrated by low AIC and BIC values and an RMSEA value below the recommended cut-off (<0.06; See Table 2). Most discrimination values were above 1.70, indicating a perfect fit of the items. The only exception was item SE6, which presented a moderate discrimination value ($a = 1.17$; See Table 2), indicating a somewhat weaker relationship of

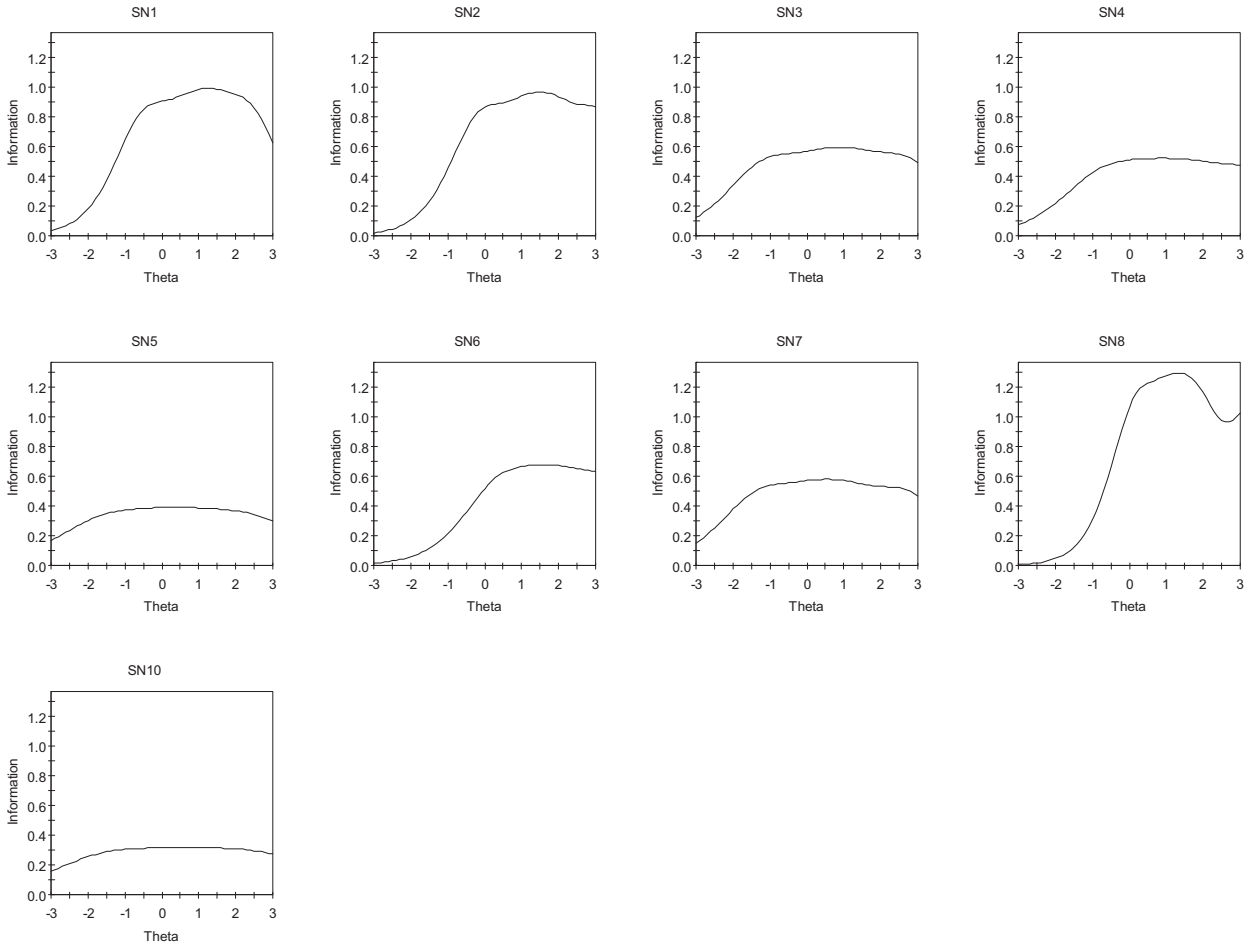


Fig. 4 Information Functions for Approval of Subjective Norms

this item with the latent construct compared to other scale items. The highest discrimination value was for SE9 and SE3, indicating a stronger relationship between these items and the latent trait. The range of thresholds along the latent trait was reasonably broad for most items (range from −7.42 to 9.22) and in the correct order, implying that most items could measure a broad range of the latent trait. LD-$\chi 2$ values for the items were acceptable (<10), supporting local independence. The IICs for both SE9 and SE3 were the highest and indicated high reliability for these two items (see Fig. 5). The lowest IIC was detected for SE6, indicating high measurement error. The removal of SE6 improved the model's fit, as demonstrated by lower goodness-of-fit statistics for the AIC, BIC, M_2, and RMSEA (see Table 1).

Reliability and Internal Consistency Results

This second set of results relates to the classical reliability and validity analyses conducted on the PDEPQ-R after the confirmatory IRT analyses.

Internal Consistency

For the sample of 400 parents, the 9-item version of the Approval of Physical Punishment scale was found to have excellent internal consistency ($\alpha = 0.92$). The 7-item version of the Approval of Non-Physical Punishment scale was found to have acceptable internal consistency ($\alpha = 0.74$). The 9-items version of the Subjective Norms scale showed good internal consistency ($\alpha = 0.82$). Finally, the 8-item version of the Parenting Self-Efficacy showed good internal consistency ($\alpha = 0.86$). For the test-retest reliability assessment using the subsample of 100 parents, correlations for the PDEPQ-R scales were very strong from the first to second administration four weeks later and ranged from 0.81 to 0.94. The highest correlation was for the Approval of Physical Punishment ($r = 0.94$, $p < 0.001$) scale, followed by Approval of Non-Physical Punishment ($r = 0.85$, $p < 0.001$), Subjective Norms ($r = 0.83$, $p < 0.001$), and Parenting Self-Efficacy ($r = 0.81$, $p < 0.001$).

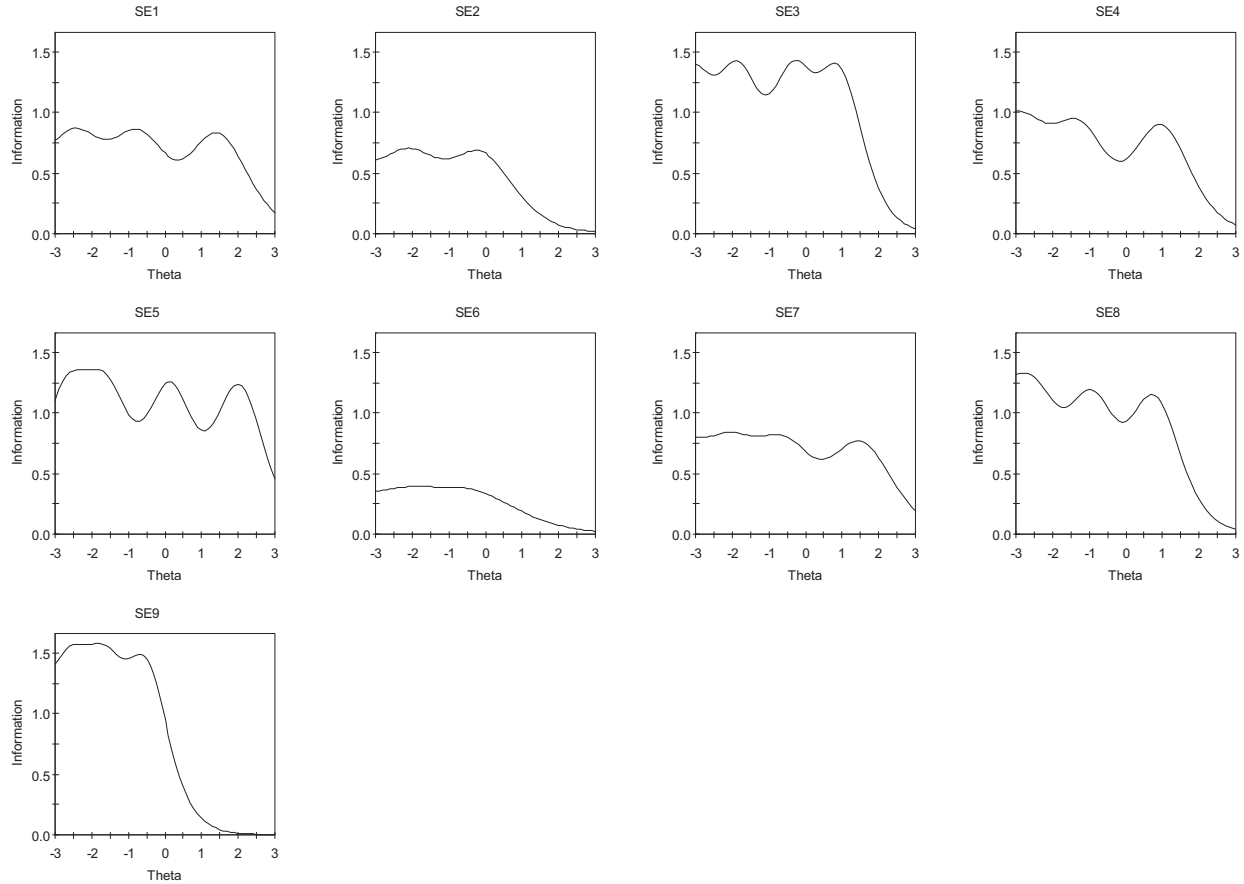


Fig. 5 Information Functions for Self-Efficacy Items

Convergent Validity

There was a robust and significant positive correlation between Attitudes Toward Spanking (ATS) and the PDEPQ-R Approval of Physical Punishment scale ($r = 0.92$; $p < 0.001$), indicating that greater parental approval of spanking was associated with greater approval of physical punishment more generally. Moreover, there was a moderate and significant positive correlation between the ATS and the Approval of Non-Physical Punishment scale ($r = 0.47$; $p < 0.001$), thereby indicating that parents who held more favourable attitudes toward spanking were more likely to approve of non-physical forms of punishment, such as non-physical punishment. In addition, there was a small positive correlation between the Child-Responsible Attributions subscale of the Parental Cognition Scale and the PDEPQ-R Subjective Norms scale ($r = 0.14$; $p = 0.004$), indicating that parents who tended to attribute challenging child behaviours to children's characteristics also had more negative general beliefs regarding normative child behaviours. However, the association between these two variables was very weak. Finally, Discipline Self-Efficacy subscale scores were significantly and positively related to Parenting Self-Efficacy scores of the PDEPQ-R ($r = 0.62$; $p < 0.001$). This finding indicates that parents who reported greater effectiveness in implementing their parenting strategies (regardless of whether they are punitive) also reported greater effectiveness in engaging in non-punitive disciplinary strategies.

Discussion

This study evaluated and modified a questionnaire designed to assess constructs that characterise some of the critical determinants of parental use of physical and non-physical punishment. The four scales of the PDEPQ-R assess parental approval of physical punishment, parental approval of non-physical (emotional) punishment, subjective norms, and parenting self-efficacy. In Study 1, the content of the original version of the scales was examined by two samples, namely professionals in child development and parenting research as well as parents. The items were examined regarding representativeness and clarity, and the two samples offered suggestions about wording changes. Following

this content validity assessment, we made changes when both participant groups highlighted the same suggestions.

Study 2 examined the psychometric characteristics of the newly revised questionnaire, PDEPQ-R, using a sample of 400 parents. A subsample of the parents was also re-assessed four weeks later to establish test-rest reliability. The validity of the constructs was examined by correlating the PDEPQ-R scales with other measures assessing similar constructs. The results provided evidence of good psychometric qualities for the PDEPQ-R. However, the analyses highlighted several problematic items that subsequently were removed. Following the removal of these items, the final version of the PDEPQ-R was confirmed. Using this final version of the PDEPQ-R questionnaire, a reliability (i.e., internal consistency and test-retest) and validity (i.e., convergent) assessment of the scales was conducted. Three of the four scales (the exception was the Approval of Non-Physical Punishment) had good internal consistency, indicating reasonably good agreement among items within the scales. Moreover, the test-retest results showed that all four PDEPQ-R scales were reliable in 100 parents over four weeks. Test–retest reliability coefficients ranged from good (.81 for Parenting Self-Efficacy) to excellent (.94 for Approval of Physical Punishment), providing strong evidence supporting the PDEPQ-R's temporal consistency and invariance over time.

Furthermore, convergent validity was generally supported based on the sample of 400 parents. In line with the first hypothesis, the Attitudes Toward Spanking (ATS) score was significantly and positively associated with both the Approval of Physical Punishment and the Approval of Non-Physical Punishment scales from the PDEPQ-R. Therefore, the Approval of Physical Punishment scale taps into aspects of parenting closely related to facets of the ATS scale; indeed, many of the items on the two scales are similar in that they assess parental approval of spanking. This close alignment between the two measures presumably accounts for their high correlation. Interestingly, ATS scores were also positively related to Approval of Non-Physical Punishment, indicating that attitudes regarding spanking are linked with attitudes toward using non-physical punishment. This finding is novel in that it is the first to indicate a link between attitudes toward using "mild" forms of physical punishment, such as spanking, and approval of non-physical punishment. Parents with more favourable attitudes toward physical punishment appear more likely to hold favourable attitudes toward punishment strategies more generally (e.g., removal of activities, time-out). These findings align with research studies that have found small to moderate correlations between parental use of physical and non-physical punishment, as measured by the Corporal Punishment and Non-Violent Discipline subscales of the Conflict Tactic Parenting Scale (CTPS; Straus et al., 2003; Rodriguez, 2010). However, these subscales measure parents' endorsement of behaviour rather than their attitudes. In this light, the PDEPQ-R contributes to existing research by providing additional evidence supporting these relationships. However, further research is needed to corroborate these findings.

In line with the second hypothesis, the Child-Responsible Attributions subscale of the Parental Cognition Scale (PCS) was positively and significantly related to the PDEPQ-R Subjective Norms scale. These findings indicate the link between parental beliefs about age-appropriate behaviours and attributions relative to the intentionality of the child (i.e., the child as the cause of challenging behaviour). Despite the significant correlation, the small magnitude of the coefficient indicates a weak alignment between the two constructs. However, it is essential to note that items on the Child-Responsible Attributions subscale tap into parents' negative attributions about their children. In contrast, the Subjective Norms scale assesses parental beliefs about childhood behaviours at different ages and, as such, is not related exclusively to their child. This difference might explain the weak relationship between these two measures.

Finally, in line with the third hypothesis, the Discipline Self-Efficacy subscale was significantly and positively related to Parenting Self-Efficacy scale of the PDEPQ-R, indicating a strong alignment between these two constructs. It is important to note that, despite the relationship between the two scales in assessing one's general sense of competence in the parenting role, the items of PDEPQ-R's Parenting Self-Efficacy scale were more specific to the use of punishment and the ability to manage challenging parenting situations constructively (i.e., without anger or the use of punishment). In contrast, the Discipline Self-Efficacy scale items focus on self-efficacy in the global sense of discipline (e.g., setting rules and overseeing decisions around parenting) with no direct reference to using punishment or managing conflict with children.

Overall, the study results indicated that the PDEPQ-R possesses good psychometric properties and constitutes a promising instrument for assessing parenting practices. Although the PDEPQ-R was designed to assess the outcomes of the PDEP programme, it can be considered an evaluation tool for any parenting programme whose aim is to reduce child physical and non-physical punishment and promote positive disciplinary strategies. Moreover, PDEPQ-R's focus on factors related to the use of child punishment (rather than a total score) makes it possible to examine in greater detail such critical factors as the approval of physical and non-physical punishment, subjective norms, and parenting self-efficacy. As such, the PDEPQ-R can be a helpful tool for advancing knowledge on the parental cognitions underlying child punishment

strategies. Furthermore, the PDEPQ-R could potentially be used in clinical settings to provide health professionals with additional information on factors influencing parents' use of various child disciplinary practices. This information can help guide clinical decisions around ways to tailor interventions to meet the specific needs of parents seeking support around child discipline.

Limitations and Future Directions

One study limitation included sample representativeness in that the sample of parents was almost exclusively composed of mothers. Only 6.2% of fathers participated in the IRT analyses, and no fathers were involved in the content validity of the PDEPQ-R based on focus groups. It is possible that the questionnaire's psychometric properties would differ depending on the parent's gender, so future research should attempt to validate (and perhaps adapt) these scales in father samples. Similarly, most parents in this study identified as White and highly educated, so additional questionnaire validity efforts with more diverse populations of parents are warranted. It should be noted, however, that an advantage of using IRT parameter estimates is that this statistical method is considered population-independent (i.e., it should provide relatively stable estimates across different populations). Second, potential weaknesses of the Subjective Norms scale indicate the need to further validate this construct to measure parental attribution to children's behaviour.

Third, this study did not use attention checks and other manipulation techniques, which could have potentially negatively impacted scale validity. In addition, other forms of psychometric evidence remain to be tested (e.g., predictive, discriminant validity), so additional research in these domains needs to be considered. Fourth, the exclusive reliance on parent self-report could represent a validity threat due to shared method variance. Therefore, expanding the current results by including observational measures of attitudes and attributions would be important in subsequent validation research efforts of the PDEPQ-R. Finally, rewording the gendered pronouns in the PDEPQ-R to be more inclusive by using gender-neutral language would be important.

Compliance with Ethical Standards

Conflict of Interest We verify that there are no potential conflicts of interest pertaining to this manuscript and that the treatment of all participants was in accordance with the ethical standards of the American Psychological Association. The study was approved by the Office of Research Ethics and Integrity of the University of Ottawa (REB # H-02-18-352). The authors received no financial support for the research, authorship and/or publication of this article.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Allport, G. W. (1935). Attitudes. In C. M. Murchison (Ed.), *Handbook of Social Psychology* (pp. 798–844). Winchester, MA: Clark University Press.

Akaike, H. (1974). A new look at the statistical model identification. *Transactions on Automatic Control*, *19*, 716–723. https://doi.org/10.1109/TAC.1974.1100705.

American Academy of Pediatrics. (2018). What's the best way to discipline my child? Retrieved from: https://www.healthychildren.org/English/family-life/family-dynamics/communication-discipline/Pages/Disciplining-Your-Child.aspx.

Ateah, C. A., & Durrant, J. E. (2005). Maternal use of corporal punishment in response to child misbehavior: Implications for child abuse prevention. *Child Abuse & Neglect*, *29*, 169–185. https://doi.org/10.1016/j.chiabu.2004.10.010.

Baker, F. (2001). The basics of item response theory. ERIC Clearinghouse on Assessment and Evaluation, University of Maryland College Park, MD.

Barnes, C. R., & Adamson-Macedo, E. N. (2007). Perceived Maternal Parenting Self-Efficacy (PMP S-E) tool: development and validation with mothers of hospitalized preterm neonates. *Journal of Advanced Nursing*, *60*(5), 550–560. https://doi.org/10.1111/j.1365-2648.2007.04445.x.

Bugental, D. B., & Johnston, C. (2000). Parental and child cognitions in the context of the family. *Annual Review of Psychology*, *51*, 315–344. https://doi.org/10.1146/annurev.psych.51.1.315.

Cai, L., Thissen, D., & du Toit, S. H. C. (2011). *IRTPRO for Windows [Computer software]*. Lincolnwood, IL: Scientific Software International.

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*, *22*, 265–289. https://doi.org/10.2307/1165285.

Coleman, P. K., & Karraker, K. H. (2000). Parenting self-efficacy among mothers of school-age children: Conceptualization, measurement, and correlates. *Family Relations*, *49*, 13–24. https://doi.org/10.1111/j.1741-3729.2000.00013.x.

Crouch, J. L., Irwin, L. M., Milner, J. S., Skowronski, J. J., Rutledge, E., & Davila, A. L. (2017). Do hostile attributions and negative affect explain the association between authoritarian beliefs and harsh parenting? *Child Abuse & Neglect*, *67*, 13–21. https://doi.org/10.1016/j.chiabu.2017.02.019.

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Public.

Dadds, M. R., & Tully, L. A. (2019). What is it to discipline a child: What should it be? A reanalysis of time-out from the perspective of child mental health, attachment, and trauma. *American Psychologist*, *74*(7), 794 https://doi.org/10.1037/amp0000449.

Durrant, J. E. (2007, 2013). *Positive discipline: What it is and how to do it*. Save the Children Sweden Southeast Asia and the Pacific.

Durrant, J. E., Plateau, D. P., Ateah, C., Stewart-Tufescu, A., Jones, A., Ly, G., Barker, L., Holden, G., Kearley, C., MacAulay, J., Peters, R. D., & Tapanya, S. (2014). Preventing punitive violence: preliminary data on the Positive Discipline in Everyday Parenting (PDEP) Program. *Canadian Journal of Community Mental Health*, *33*, 109–125. https://doi.org/10.7870/cjcmh-2014-018.

George, D., & Mallery, P. (2003). SPSS for Windows step by step: A simple guide and reference. 11.0 update (4th ed.). Boston: Allyn & Bacon.

Gershoff, E. T. (2002). Corporal punishment by parents and associated child behaviors and experiences: A meta-analytic and theoretical review. *Psychological Bulletin*, *128*, 539–579. https://doi.org/10.1037//0033-2909.128.4.539.

Gershoff, E. T., Grogan-Kaylor, A., Lansford, J. E., Chang, L., Zelli, A., Deater-Deckard, K., & Dodge, K. A. (2010). Parent discipline practices in an international sample: Associations with child behaviors and moderation by perceived normativeness. *Child development*, *81*(2), 487–502. https://doi.org/10.1111/j.1467-8624.2009.01409.x.

Gershoff, E. T., & Grogan-Kaylor, A. (2016). Spanking and child outcomes: Old controversies and new meta-analyses. *Journal of Family Psychology*. https://doi.org/10.1037/fam0000191.

Fang, Y., Boelens, M., Windhorst, D. A., Raat, H., & van Grieken, A. (2021). Factors associated with parenting self-efficacy: A systematic review. *Journal of Advanced Nursing*, *77*(6), 2641–2661. https://doi.org/10.1111/jan.14767.

Gittins, C. B., & Hunt, C. (2019). Parental behavioural control in adolescence: How does it affect self-esteem and self-criticism? *Journal of Adolescence*, *73*, 26–35. https://doi.org/10.1016/j.adolescence.2019.03.004.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Hamovitch, E. K., Acri, M. C., & Bornheimer, L. A. (2019). An analysis of the relationship between parenting self-efficacy, the quality of parenting, and parental and child emotional health. *Journal of Family Social Work*, *22*(4-5), 337–351. https://doi.org/10.1080/10522158.2019.1635939.

Holden, G. W., Grogan-Kaylor, A., Durrant, J. E., & Gershoff, E. T. (2017). Researchers deserve a better critique: Response to Larzelere, Gunnoe, Roberts, and Ferguson. *Marriage & Family Review*, *53*(5), 465–490. https://doi.org/10.1080/01494929.2017.1308899.

Holden, G. W. (2020). Why do parents hit their children? From cultural to unconscious determinants. *The Psychoanalytic Study of the Child*, *73*(1), 10–29. https://doi.org/10.1080/00797308.2020.1690858.

Holden, G. W., & Buck, M. J. (2002). Parental attitudes toward child rearing. In M. H. Bornstein (Ed.), *Handbook of Parenting*. 2nd ed. (pp. 537–562). Mahwah, NJ: Erlbaum.

Holden, G. W., Coleman, S. M., & Schmidt, K. L. (1995). Why 3-year-old children get spanked: Parent and child determinants as reported by college-educated mothers. *Merrill-Palmer Quarterly*, *41*, 431–452. https://www.jstor.org/stable/23087935.

Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. https://doi.org/10.1080/10705519909540118.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory*. Homewood, IL: Dow Jones-Irwin.

Kaminski, J. W., & Claussen, A. H. (2017). Evidence base update for psychosocial treatments for disruptive behaviors in children. *Journal of Clinical Child & Adolescent Psychology*, *46*, 477–499. https://doi.org/10.1080/15374416.2017.1310044.

Khoury-Kassabri, M., Attar-Schwartz, S., & Zur, H. (2014). Understanding the mediating role of corporal punishment in the association between maternal stress, efficacy, co-parenting and children's adjustment difficulties among Arab mothers. *Child Abuse & Neglect*, *38*(6), 1073–1082.

Knight, R. M., Albright, J., Deling, L., Dore-Stites, D., & Drayton, A. K. (2020). Longitudinal relationship between time-out and child emotional and behavioral functioning. *Journal of Developmental & Behavioral Pediatrics*, *41*(1), 31–37. https://doi.org/10.1097/DBP.0000000000000725.

Krueger, R. A. (2014). *Focus groups: A practical guide for applied research*. Sage publications.

Larzelere, R. E., Gunnoe, M. L., Roberts, M. W., & Ferguson, C. J. (2017). Children and parents deserve better parental discipline research: Critiquing the evidence for exclusively "positive" parenting. *Marriage & Family Review*, *53*(1), 24–35. https://doi.org/10.1080/01494929.2016.1145613.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

Maydeu-Olivares, A., & Joe, H. (2005). Limited-and full-information estimation and goodness-of-fit testing in 2 n contingency tables: A unified framework. *Journal of the American Statistical Association*, *100*(471), 1009–1020. https://doi.org/10.1198/016214504000002069.

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, *71*(4), 713 https://doi.org/10.1007/s11336-005-1295-9.

Marsteller, J. A., Hsu, Y. J., Chan, K. S., & Lubomski, L. H. (2017). Assessing content validity and user perspectives on the Team Check-up Tool: Expert survey and user focus groups. *BMJ Quality & Safety*, *26*(4), 288–295. https://doi.org/10.1136/bmjqs-2015-004635.

Milner, J. S. (2003). Social information processing in high-risk and physically abusive parents. *Child Abuse & Neglect*, *27*, 7–20. https://doi.org/10.1016/S0145-2134(02)00506-9.

Mueller, R. O. (1997). Structural equation modeling: Back to basics. *Structural Equation Modeling: A Multidisciplinary Journal*, *4*(4), 353–369.

Perron, J. L., Lee, C. M., LaRoche, K. J., Ateah, C., Clément, M. È., & Chan, K. (2014). Child and parent characteristics associated with Canadian parents' reports of spanking. *Canadian Journal of Community Mental Health*, *33*, 31–45. https://doi.org/10.7870/cjcmh-2014-014.

Querido, J. G., Bearss, K., & Eyberg, S. M. (2002). Parent/child interaction therapy. In F. W. Kaslow & T. Patterson (Eds.), *Comprehensive handbook of psychotherapy: Cognitive-behavioral approaches* (Vol. 2, pp. 91–113). Hoboken, NJ: Wiley.

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor test: Results and implications. *Journal of Educational Statistics*, *4*, 207–230. https://doi.org/10.2307/1164671.

Reise, S. P., & Henson, J. M. (2003). A discussion of modern versus traditional psychometrics as applied to personality assessment scales. *Journal of Personality Assessment*, *81*, 93–103. https://doi.org/10.1207/S15327752JPA8102_01.

Rodriguez, C. M. (2010). Parent–child aggression: Association with child abuse potential and parenting styles. *Violence and Victims*, *25*(6), 728–741. https://doi.org/10.1891/0886-6708.25.6.728.

Rodriguez, C. M., & Wittig, S. M. (2019). Predicting child problem behavior and maternal/paternal parent-child aggression: identifying early prevention targets. *Journal of Applied Developmental Psychology*, *60*, 76–86. https://doi.org/10.1016/j.appdev.2018.11.001.

Rominov, H., Giallo, R., & Whelan, T. A. (2016). Fathers' postnatal distress, parenting self-efficacy, later parenting behavior, and children's emotional-behavioral functioning: A longitudinal study. *Journal of Family Psychology*, *30*(8), 907 https://doi.org/10.1037/fam0000216.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *34*, 100–114. https://doi.org/10.1002/j.2333-8504.1968.tb00153.x.

Sanders, M. R., Kirby, J. N., Tellegen, C. L., & Day, J. J. (2014). The Triple P-Positive Parenting Program: A systematic review and metaanalysis of a multi-level system of parenting support. *Clinical Psychology Review*, *34*, 337–357. https://doi.org/10.1016/j.cpr.2014.04.003.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464. https://www.jstor.org/stable/2958889.

Snarr, J. D., Slep, A. M. S., & Grande, V. P. (2009). Validation of a new self-report measure of parental attributions. *Psychological Assessment*, *21*, 390–401. https://doi.org/10.1037/a0016331.

Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for numbers of common factors*. Iowa City, IA: Paper presented at the annual meeting of the Psychometric Society.

Straus, M. A., Hamby, S. L., & Warren, W. L. (2003). *The conflict tactics scales handbook*. Los Angeles, CA: Western Psychological Services.

Sturge-Apple, M. L., Li, Z., Martin, M. J., Jones-Gordils, H. R., & Davies, P. T. (2020). Mothers' and fathers' self-regulation capacity, dysfunctional attributions and hostile parenting during early adolescence: A process-oriented approach. *Development and Psychopathology*, *32*(1), 229–241.

Vittrup, B., Holden, G. W., & Buck, J. (2006). Attitudes predict the use of physical punishment: A prospective study of the emergence of disciplinary practices. *Pediatrics*, *117*, 2055–2064. https://doi.org/10.1542/peds.2005-2204.

Webster-Stratton, C. (2001). The incredible years: Parents, teachers, and children training series. *Residential Treatment for Children & Youth*, *18*, 31–45. https://doi.org/10.1300/J007v18n03_04.

Wittkowski, A., Garrett, C., & Calam, R., et al. (2017). Self-report measures of parental self-efficacy: A systematic review of the current literature. *Journal of Child Family Studies*, *26*, 2960–2978. https://doi.org/10.1007/s10826-017-0830-5.