Child Protection and Practice 8 (2026) 100267

Contents lists available at ScienceDirect

Child Protection and Practice

journal homepage: www.sciencedirect.com/journal/child-protection-and-practice

Adapting and validating the client engagement scale for the Danish child and family welfare context: A Rasch-based construct validity study

Signe Boe Rayce [a,*], Anne Marie Villumsen [a], Maiken Pontoppidan [a], Tine Nielsen [b]

[a] VIVE – the Danish Center for Social Science Research, Herluf Trollesgade 11, DK 1052, København K, Denmark
[b] UCL University College, Department of Applied Research in Education and Social Sciences, Niels Bohrs Allé 1, DK 5230, Odense, Denmark

ARTICLE INFO

Keywords:
Client engagement
Child and family welfare system
Psychometric properties
Rasch models
Parents

ABSTRACT

Background: Client engagement encompasses family involvement in child protection casework and the collaborative relationship between case workers and families. The Client Engagement Scale was developed to assess client engagement within child protective systems.
Objective: Building on Yatchmenoff's conceptual framework, this study aimed to translate and conduct an initial psychometric validation of a client engagement measure specifically within Denmark's child and family welfare system.
Participants and setting: Parents with active cases in the Danish Child and Family Welfare system completed 210 online survey questionnaires.
Methods: The original pool of 37 items was translated and adapted to the Danish context. Construct validity for five subscales of client engagement was examined using Rasch models (RM) and graphical loglinear Rasch models (GLLRM) with a focus on differential item functioning and local dependency of items.
Results: After item elimination, a 5-item expectancy subscale fitted the pure RM. A 6-item receptivity, 4-item mistrust, and a 3-item working relationship subscale presented some differential functioning and/or local dependency, but to a degree which could be accounted for using GLLRM. The investment subscale did not fit the RM nor a GLLRM. The resulting scales demonstrated high to excellent reliability, but poor targeting to the study population.
Conclusion: Essentially valid subscales resulted for four out of five aspects of client engagement within Denmark's Child and Family Welfare system. While these subscales show promise for statistical use, future research should examine their psychometric properties in larger and more diverse samples before attempting to use them at an individual level in an everyday child and family welfare practice.

1. Introduction

Establishing trusting and collaborative relationships between service users and their caseworkers in statutory social services is both critical and inherently challenging, as power imbalances can hinder genuine collaboration (Bekaert et al., 2021; Samsonsen & Willumsen, 2015; Tembo & Studsrød, 2019; Thrana & Fauske, 2014). Moreover, defining and conceptualizing such a relationship is complex, particularly in non-voluntary contexts. The concept of engagement represents an attempt to operationalize this complex and dynamic interaction.

Engagement, as conceptualized by Trotter (Trotter, 2002), encompasses family involvement in child protection casework and the collaborative relationship between social workers and families in mandated contexts. Research demonstrates that engagement is foundational to effective interventions, as it is linked to improved family experiences (Cheng & Lo, 2016; Estefan et al., 2012; Gladstone et al., 2014; Kemp et al., 2014; Xu et al., 2017) and better service outcomes (Cheng & Lo, 2016; Damiani-Taraba et al., 2017; Fernandez, 2004; Gladstone et al., 2012, 2014; Lee & Ayón, 2004). Research on parental engagement in social services closely align with broader studies on parental experiences in mandated collaborations, both consistently emphasizing the importance of the parent-social worker relationship. Parents highlight qualities such as kindness (Studsrød et al., 2014), care (Maiter et al., 2006; Thrana & Fauske, 2014), and empathy (Bouma et al., 2020;

This article is part of a special issue entitled: Child maltreatment data in research and practice published in Child Protection and Practice.

* Corresponding author.
E-mail addresses: sbo@vive.dk (S.B. Rayce), amav@vive.dk (A.M. Villumsen), mpo@vive.dk (M. Pontoppidan), tini@ucl.dk (T. Nielsen).

https://doi.org/10.1016/j.chipro.2025.100267
Received 2 May 2025; Received in revised form 8 September 2025; Accepted 3 December 2025
Available online 5 December 2025
2950-1938/© 2025 The Authors. Published by Elsevier Inc. on behalf of International Society for Prevention of Child Abuse and Neglect. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Brandon et al., 2019; Maiter et al., 2006; Merkel-Holguin et al., 2015; Moesby-Jensen & Moesby-Jensen, 2019), and emphasize the importance of effective communication. Experiences of disempowerment can arise when information is poorly handled – whether the information is unexpectedly disclosed (Coady et al., 2013), withheld (Studsrød et al., 2014), or denied (Smithson & Gibson, 2017). In addition, stronger collaboration between caseworkers and family is positively associated with improved compliance, which in turn significantly reduces reports of child maltreatment and the need for out-of-home placements (Littell, 2001).

Accurately assessing client engagement is critical for understanding its influence on intervention outcomes and refining practices accordingly. However, developing valid measurement instruments to assess engagement in a collaborative relationship between client and caseworker is challenging. In response to the need for valid instruments, Yatchmenoff (Yatchmenoff, 2005) developed the Client Engagement Scale (CES) – an instrument specifically designed to address engagement in non-voluntary social work settings. The CES, or parts of it, has been utilized in multiple studies across the United States and Canada to examine various aspects of engagement within child protection systems. Research using the CES has explored how family-related factors predict engagement in child welfare services (Mirick, 2014a) and the relationship between client reactance and engagement in these services (Mirick, 2014b). Other CES studies have found associations between strength-based caseworker practice and parents' willingness to engage in child welfare services (Kemp et al., 2014). Client engagement has also been positively linked with client satisfaction with case outcomes and caseworker skills, respectively(Damiani-Taraba et al., 2017; Gladstone et al., 2012, 2014). This emphasizes the importance of specific caseworker skills in fostering collaborative relationships. While these studies contribute valuable insights into client engagement within North American social service systems, none have focused on validating the CES in other languages, other national settings or differing types of social services systems.

While the child social service systems in Canada and the United States are typically framed as child protection systems, which typically intervene only after substantiated cases of harm or maltreatment (Gilbert, 1997; Gilbert et al., 2011; Khoo et al., 2003; Križ & Skivenes, 2013), the Nordic social service systems operate within a broader welfare state framework. The Danish social service system is characterized by universalism, high levels of public spending and taxation, and extensive state involvement in welfare provision. This model fosters greater equality in access to resources such as education, healthcare, and social services (Greve, 2007). Both universal and targeted services are integrated within a child-centered approach that prioritizes family preservation and close collaboration with those in need (Pösö et al., 2014). Preventive measures are a core focus, and the threshold for intervention is relatively low compared to child protection-oriented systems in North America.

Another key distinction between the Danish and North American societal models and their respective social service systems lies in the level of risk faced by children and families. In Nordic countries, structural features such as income redistribution, poverty reduction, and welfare services reduce risk exposure compared to those observed in Canada and the United States (Ólafsson, 2013). Such differences may influence service delivery and outcomes (Benbenishty et al., 2015). This may in turn, influence the dynamics of client engagement. This underscores the importance of validating measures of client engagement in other national settings and differing types of social services systems.

The validity of the CES has been investigated through confirmatory factor analysis (Yatchmenoff, 2005), however, this did not include tests of differential item functioning or invariance, i.e. whether items function differently across subgroups of the population. No other studies have examined the items and how they function by employing modern item analysis, i.e. by Item Response Theory (IRT) models or Rasch models.

Building on Yatchmenoff's (Yatchmenoff, 2005) foundational work, this study translates and conducts initial psychometric item analyses of a client engagement measure within Denmark's child and family welfare system. Employing Rasch models, the study examines the psychometric properties of scales of five aspects of client engagement, with a particular focus on item functioning: Which items can be included in the Danish scales? Do items exhibit differential item functioning across subgroups? Are there instances of locally dependent items?

By addressing these questions, this study aims to contribute to the development of construct valid tools for assessing client engagement in the Danish Child and Family Welfare system. In doing so, it aims to advance the conceptual understanding of client engagement and improve its practical application in non-voluntary casework.

2. Methods

2.1. Sample

The study sample comprises parents with active child cases in the Danish Child and Family Welfare system. Data was collected through an online survey among parents recruited from five municipalities. In four of the municipalities, a total of 48 caseworkers recruited parents they were working with through child and family welfare services (CWS). In the fifth municipality, parents were recruited by a CWS secretary. Upon obtaining consent, the research team received the families' contact details.

A total of 578 questionnaires were distributed to parents by email, each containing a link to the online questionnaire. To maximize participation, automated reminders were sent after one week, followed by manual follow-ups as needed.

The final sample consists of 210 completed questionnaires, corresponding to a response rate of 36 %. These 210 data cases were provided by 182 parents, as 22 parents completed the questionnaire for more than one child (2–4 children). These additional responses were retained in the sample as they refer to distinct child-specific cases and are therefore considered independent at the child level, Inspection of data supported this assumption, i.e. parents with more than one child provided different responses to the questionnaires across their children. At the parent level, the responses are, however, not independent.

The study was approved by the Danish Center for Social Science Research's Research Ethics Committee. All parents received written information about the survey, including details on the voluntary nature of participation, strict confidentiality of their data, and compliance with General Data Protection Regulation for data storage and analysis.

Table 1 shows characteristics of the parents' contact with the child and welfare system and the socio-demographic characteristics of the sample.

2.2. Instrument

The Client Engagement Scale (Yatchmenoff, 2005) is an instrument developed to assess client engagement in child protection services. The CES was initially developed from a pool of 37 items derived from qualitative interviews with supervisors, caseworkers, frontline practitioners, and an advisory group, all of whom reflected on the nature of client engagement within child protective services (Appendix A). The 37 original items were designed to measure five sub-constructs of client engagement: *receptivity* (openness to receiving help and characterized by recognition of problems and the need for help), *working relationship* (relationship with worker characterized by good communication and a sense of reciprocity), *expectancy* (perception of benefit with a sense of being helped or an expectation of getting help from the agency and that things will be better), *investment* (commitment to the process through active participation in services and planning of this as well as active help seeking) and *mistrust* (a belief of the agency or workers being manipulative with an intent to harm the client). Following two rounds of

Table 1
Characteristics of the study sample (n = 210).

Characteristics	n	%
Parent gender		
Female	151	71.9
Male	59	28.1
Parent education		
Medium or long-cycle higher education[a]	98	46.7
Vocational or short-cycle higher education[b]	59	28.1
Secondary education or less	53	25.2
Language at home		
Always Danish	193	91.9
Not always Danish	17	8.1
Child age		
0–5 years	29	13.8
6–10 years	41	19.5
11–13 years	42	20.0
14–17 years	98	46.7
Case length		
≤1 year	36	17.1
>1 year	174	82.9
Previous child case		
No previous cases	76	36.2
Previous cases with either the same or another child	134	63.8
Parent age		
21–30 years	22	10.5
31–40 years	79	37.6
41–50	80	38.1
>50 years	29	13.8

[a] Medium-cycle education: professional bachelor education (e.g. nurse and teacher); Long-cycle education: master- or Ph.D.-degree.

[b] Vocational or short-cycle higher education: Vocational education: commercial and technical vocational education (e.g. retail, carpenter, social- and healthcare assistant); Short-cycle higher education: vocational academy education of typically 2 years duration (e.g. pharmacy assistant, transport logistician, construction technologist).

confirmatory factor analysis, first 11 and then further 7 items were eliminated – resulting in a 19-item scale. In the final model, the original subscales *expectancy* and *investment* were combined into a single subscale, *buy-in*, and items with similar content were eliminated while ensuring that each subscale retained at least three items.

The 19-item CES (Yatchmenoff, 2005) comprises four subscales assessing different aspects of client engagement: *receptivity* (4 items), *buy-in* (8 items), *working relationship* (4 items) and *mistrust* (3 items). Each item statement is rated on a 5-point Likert response scale (1: strongly disagree, 2: disagree, 3: not sure, 4: agree to 5: strongly agree), and negatively worded items are subsequently reverse-scored (see Appendix A). Higher sum scores on the subscales indicate greater client engagement. The dimensionality of the 19-item CES was tested through confirmatory factor analysis in a sample of 287 American mothers involved with child protective services, 69 % of whom resided in the metropolitan area of Portland. A majority of the mothers had experienced child removal (69 %), substance abuse (56 %), or domestic violence (55 %) (Yatchmenoff, 2005). Cronbach's α for this sample were 0.81 (*receptivity*), 0.88 (*mistrust* and *working relationship*) and 0.91 (*buy-in*). In a Canadian sample of 165 parents with a child placed in family-based care, Cronbach's α of the *receptivity* subscale was 0.72 and 0.86 for the *working relationship* subscale (Charest-Belzile et al., 2020). Tests for differential item functioning, invariance and local dependency were, however, not conducted.

Given the differences between the American child protective services system and the Danish Child and Family Welfare context, we hypothesized that some items that were removed in the American version could hold relevance in a Danish context. Thus, we decided to include all 37 original CES items in our study.

Translation: The 37 items were translated into Danish following established methods for translating psychological and educational tests across cultural contexts (Merenda, 2005; Oakland, 2005; Sireci, 2005). Two native Danish-speaking subject matter experts independently translated the items. Minor adaptations were made to ensure cultural and linguistic alignment with the Danish context in collaboration with a third researcher. Subsequently, a native English speaker conducted the back-translation. This back-translated version was compared to the original English items and found to have satisfactory equivalence. To further evaluate the appropriateness of the Danish translation, the items were reviewed by a colleague with subject matter expertise, three leaders of the CWS in two municipalities, and three parents with active CWS cases. This process led to adaptations of items 6, 7 17, 24, 25, 28 and 37 (see Appendix A). Items 25, 28 and 37 were reworded from negative to positive phrasing to reduce the potential for misunderstandings caused by negations (Johnson et al., 2004). The remaining adaptations primarily addressed linguistic clarity. Additionally, we revised response category 3. Originally labeled 'not sure' in the English version, it was modified to "neither agree nor disagree" in the Danish version, as "not sure" can have multiple interpretations and is not a true midpoint on the response scale.

The Danish adaptation of the original 37 items, alongside their original English wording, is presented in Appendix A. Details on the adaptions can be obtained from the first author.

2.3. Item analysis using Rasch measurement models

In the present study we used Rasch models (RM) (Rasch, 1960) and Graphical Loglinear Rasch models (GLLRMs) (Kreiner & Christensen, 2002, 2004, 2007). Rasch models are a family of Item Response Theory (IRT) models comprising models for both dichotomous and polytomous items (Fischer & Molenaar, 1995). The Rasch model was first developed for dichotomous items (Rasch, 1960), and has since been generalized to polytomous ordinal items in different parameterisations (Mesbah & Kreiner, 2013), for example the partial credit model (Masters, 1982) used in this study. The dichotomous and polytomous Rasch models adhere to the same requirements for fit, and thus we simply use the term 'RM' for Rasch model (Kreiner, 2013; Mesbah & Kreiner, 2013).

In order for a set of responses to items to fit a RM, five basic requirements have to be met. 1): Unidimensionality: The items in a scale should measure one underlying latent variable only, e.g. *receptivity*. 2) Monotonicity: The expected item scores should increase as the values of the latent variable increases. This means that when e.g. *receptivity* increases the expected scores of each of the items in the scale also increases; 3) Local independence (or no local dependence, no LD): item responses must be conditionally independent given the latent variable meaning that only the latent variable (e.g. *receptivity*) affects the responses of the items, not the items mutually; 4) Absence of differential item functioning (no DIF): item responses and relevant background (i.e. exogenous) variables (e.g. parent gender or child age) must be conditionally independent given the latent variable, i.e. responses to an item should depend on only the *level of* e.g. *receptivity* and not on subgroup characteristic; 5) Homogeneity: The rank order of item parameters (e.g. item 'difficulties') should be consistent across individuals regardless of their level on the latent variable, i.e. the 'easiest' and most 'difficult' item to endorse should be the same for all individuals irrespective of *degree of* e.g. *receptivity*. Fulfillment of the first four requirements provide criterion-related construct validity according to Rosenbaum's (Rosenbaum, 1989) definition. Fulfillment of all five requirements provides a summed scale scores that are sufficient statistics for the person parameter estimates, meaning that no additional information besides the total score of the scale is necessary to assess the level of the construct measured, e.g. receptivity (Rasch, 1960). Sufficiency is a property that only scales fitting a RM hold, and it is an attractive property for scales intended for use by practitioners, as they can then use the sum score directly (Kreiner, 2007; Rasch, 1960).

It is not uncommon for scales to have locally dependent items and/or items functioning differentially. In such cases one strategy is to remove items presenting DIF or LD. However, a possible consequence of removal of otherwise content valid items would be a reduced reliability. Thus,

when working with brief scales, we prefer to avoid elimination of items, if at all possible. Graphical loglinear Rasch models (GLLRMs), which is an extension of the RM, offer a solution to this by including interaction terms for uniform LD and DIF in the model (Kreiner & Christensen, 2007). DIF is uniform when the statistical relationship between the item responses and group is constant for all levels of a matching variable (Hanson, 1998). Similarly, we can interpret uniform LD as the situation where the statistical relationship between responses to two items is constant for all levels of a matching variable. If scales include uniform LD items by using GLLRM, the sum score is still a sufficient statistic for the person parameter estimates. If a scale includes uniform DIF, the sum score is no longer a sufficient statistic for the person parameter estimates for all persons, as additional information on subgroup membership is needed to assess their level on the measured construct. However, we can equate the sum scores for DIF by utilizing the person parameter estimates in each of the DIF subgroups to bring them back on the same scale, and thus obtain unbiased sum scores for subsequent statistical analysis (Kreiner, 2007; Kreiner & Nielsen, 2023).

2.4. Exogeneous variables

Previous research has found socio-demographic factors and long histories with service systems to be associated with parent engagement (Kemp et al., 2009). In order to test for differential item functioning and measurement invariance, we included case length, any previous child cases, child age, parental gender and education and language spoken at home as exogenous variables in the analyses.

To obtain groups for DIF analyses the exogeneous variables were categorized in the following ways: child age (0–5 years; 6–10 years; 11–13 years; 14–17 years), parental gender (women vs. men because the number of parents identifying as non-binary was too small to be included in the DIF analyses), parental education (medium- or long-cycle higher education; vocational or short-cycle higher education; secondary education or less), language spoken in the home (always Danish vs. not always Danish), case length of the current child case ($\leq$1 year vs. >1year), previous child cases (No previous cases vs. previous cases with either the same or another child).

2.5. Statistical analyses

The same analytical strategy was used for each of the five subscales. First, the full set of original items were tested against the requirements of the RM using the following test statistics: 1) conditional likelihood ratio (CLR) tests of overall homogeneity to compare item parameters across low and high scoring groups, and of invariance by comparing the full set of item parameters in subgroups defined by exogeneous variables, one exogenous variable at a time (Andersen, 1973). 2) item fit statistics comparing the observed and expected correlations under the model between single items and the restscore (i.e. the total score without the item in question) (Christensen, 2013; Kreiner, 2011). 3) The local independence of items and absence of DIF was tested using Kelderman's (Kelderman, 1984) likelihood-ratio test, and if evidence of LD or DIF was discovered the magnitude of these were informed by partial Goodman-Kruskal gamma coefficients (Kreiner & Christensen, 2004).

Second, if evidence of LD and/or DIF was found, the next step was to define a GLLRM, which included LD and/or DIF interaction terms, and test whether there was fit to this model using the same fit statistics as for the RM (c.f. the above). When fit to a GLLRM could not be reached, the most problematic item was removed from the subscale based on both psychometric results (with strong DIF and LD as well as evidence against item fit being considered the most serious problems) and conceptual considerations (approach taken from (Nielsen & Kreiner, 2013)). This was followed by a new iterative process until fit to a RM or a GLLRM was reached for a set of items.

When a final model (either RM or GLLRM) was reached for a subscale, targeting and reliability was evaluated. Reliability was calculated using Hamon and Mesbah's (Hamon & Mesbah, 2002) Monte Carlo method, because this method takes any LD between items into account and adjusts reliability accordingly. Targeting of theta was assessed by 1) the test information target index (the mean test information divided by the maximum test information for theta) and 2) and the root mean squared error (RMSE) target index (the minimum standard error of measurement (SEM) divide by the mean SEM for theta) (Kreiner & Christensen, 2013). The target and SEM of the observed score was also calculated. Finally, to provide a graphical illustration of targeting and test information, we plotted so-called item maps showing the distribution of the person locations against the item thresholds along the latent scale, and included the test information curve and the point of maximum information (i.e. the test target).

We used the Benjamini-Hochberg procedure to adjust for false discovery rate (FDR) due to multiple testing (Benjamini & Hochberg, 1995). As recommend by Cox et al. (Cox et al., 1977), we did not use a deterministic criterion of 5 % critical limits for p-values. Instead, we distinguished between weak and moderate evidence against the model ($p > 0.01$) and stronger evidence ($p < 0,01$).

All analyses were conducted in the DIGRAM software package (Kreiner, 2003; Kreiner & Nielsen, 2013, 2023). Item maps were created with R using code developed by Pedro Ribeiro Santiago.

3. Results

The initial analyses, including all original items in the respective subscales, rejected fit to a pure RM in all five cases (Appendix B, Table B.1-6). Neither was is possible to establish fit to a GLLRM for any of the five subscales when all items were included. Thus, it was necessary to eliminate items that did not meet the model requirements in various ways from all scales. The final scales are described below along with the reasons for eliminating items.

In the **expectancy subscale** the most serious issues were found with item 17: *'I feel resigned or discouraged about how my case with CWS is going to turn out'* and item 32: *'CWS is not really helping*' in the form of moderate to strong LD with other items. Thus, these items were eliminated to reach a fitting model. After eliminating of these two items a 5-item scale fitted the pure RM (Fig. 1a and Appendix A). No evidence against global homogeneity or invariance relative to the six exogenous variables was found (Table 2) and there was no evidence against fit of the individual items (Appendix C, Table C.1). The score ranged from 5 to 25 with higher scores indicating a higher degree of expectancy.

In the **receptivity subscale** the most serious issues were found with item 34: *'I know I need to make some changes in my life to help my kids'* in form of several instances of LD with other items, as well as weak evidence against fit of this item. Therefore, item 34 was eliminated to reach a fitting model. Hereafter, a 6-item scale (Appendix A) fitted a relatively simple GLLRM with two instances of DIF (items 2 and 24) relative to having a previous case in the child & protection welfare system (Fig. 1b). Thus, parents with a previous case were systematically more prone to agree with the statements of item 2 *'I realize I need some help to make sure my kids have what they need*' ($\gamma = 0.40$, $p < 0.001$) and item 24: *'There were definitely some problems in my family that CWS/my caseworker saw (when the case was initiated'* ($\gamma = 0.54$, $p = 0.001$) than parents with no previous case, no matter their level of receptivity. There was no evidence of LD between items, no evidence against global homogeneity or invariance (Table 2) and no evidence against fit of the items (Appendix C, Table C.2) after these DIF interaction terms were included in the GLLRM. The score ranged from 6 to 30 with higher scores indicating a higher degree of receptivity.

In the **mistrust subscale** the most serious issues were found with item 11: *'I don't say more than I have to when anybody from the agency is around.'* and item 23: *'I really watch my back when it comes to CWS.'* in the form of several instances of LD, including negative LD that could not be explained conceptually. Both items were eliminated to reach a fitting model. The remaining 4-item scale (Appendix A) fitted a GLLRM with

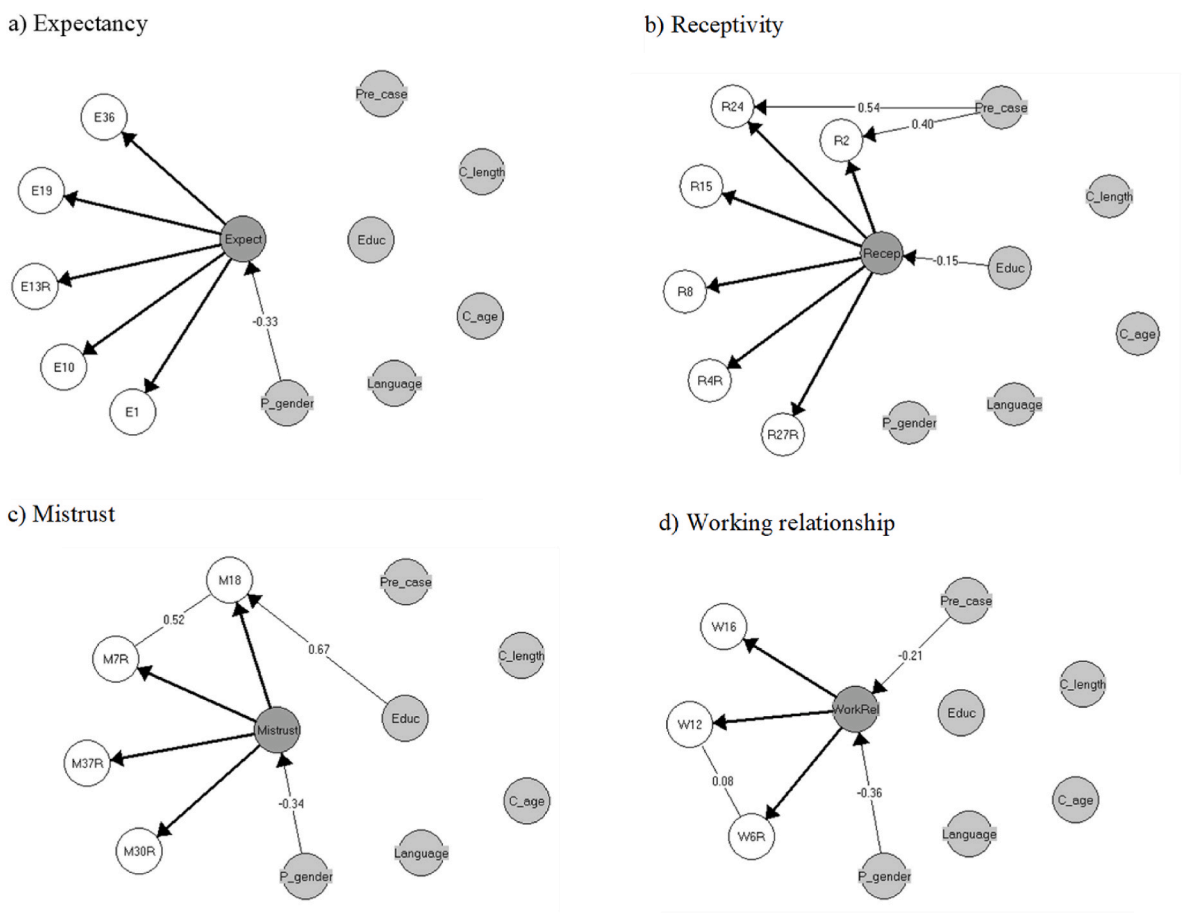


Fig. 1. Final RM or GLLRMs for the four subscales
Note: γ-correlations are partial Goodman and Kruskal's rank correlations for ordinal data. P_gender = parental gender, Educ = parental education, C_age = child age, Language = language spoken at home, C_length = Case length, Pre_case = Previous case in the child and family welfare system.

Table 2
Global Test of homogeneity and invariance for the four subscales under the final Rasch model and Graphical loglinear models.

Test of fit	Expectancy (RM)			Receptivity (GLLRM)			Working relationship (GLLRM)			Mistrust (GLLRM)		
	CLR	df	*p*	CLR	df	*p*	CLR	df	*p*	CLR	df	*p*
Global homogeneity	18.7	19	0.474	22.7	30	0.829	12.0	22	0.957	23.8	33	0.880
Invariance relative to												
Child age	53.3	57	0.616	98.8	90	0.246	78.1	66	0.146	94.7	99	0.603
Case length	12.5	19	0.862	29.0	30	0.519	21.2	22	0.506	31.3	33	0.554
Parental educational level	52.6	38	0.058	87.1	60	0.013	70.7	44	0.007	53.5	42	0.110
Language at home	29.1	19	0.065	39.0	30	0.127	32.4	22	0.070	43.2	33	0.111
Previous case	19.9	19	0.399	15.3	15	0.427	31.8	22	0.082	34.9	33	0.376
Parental gender	24.0	19	0.198	31.6	30	0.388	26.8	22	0.219	27.1	33	0.756

one instance of LD (item 7: *'Anything I say they're going to turn around'* and item 18:' *I feel like I can trust CWS to be fair and to see my side of things.*') ($\gamma = 0.52$, $p = 0.001$) and one instance of DIF (item 18) relative to parental education ($\gamma = 0.67$, $p < 0.001$). Thus, the shorter the education parents had the more prone they were to agree 'that they can trust the CWS to be fair and to see their side of things' than parents with longer educations (Fig. 1c), no matter their level of mistrust. There was no evidence against global homogeneity or invariance (Table 2) and no evidence against fit of the items (Appendix C, Table C.3) after the DIF interaction terms and LD were included in the GLLRM. The score ranged from 4 to 20 with higher scores indicating a lower degree of mistrust.

In the **working relationship subscale** four items were eliminated to reach a fitting model: Item 25 *'I have regular contact with my caseworker.'* was eliminated due to several instances of strong positive but also negative DIF indicating that frequent contact with one's caseworker may not indicate a positive working relationship. Item 28 *'My caseworker understands my reality.'* was involved in several instances of strong and positive LD indicating some degree of redundancy in items. Additionally, strong evidence against item fit was found for both items 25 and 28. After removal of items 25 and 28 both, item 21: *'I feel like I can tell my side of the story and be heard.'* and item 31: *'If I need to talk to my worker, I just call, and we talk.'* were involved in strong (item 21) or negative (item 31) instances of LD. After eliminating these four items, the 3-item scale (Appendix A) fitted a GLLRM with one instance of LD between items 6: *'It's hard for me to work with my caseworker.'* and 12: *'I think my caseworker and I respect each other.'* This LD was, however, weak ($\gamma = 0.08$, p

< 0.001) and thus without practical consequence (Fig. 1d). There was no evidence against global homogeneity or invariance (Table 2) and no evidence against fit of the items (Appendix C, Table C.4) after the LD interaction term was included in the GLLRM. The score ranged from 3 to 15 with higher scores indicating a better perception of working relationship.

Despite rigorous analyses following the strategy described in the section on statistical analyses above, it was not possible to establish fit to a GLLRM for any items in the **investment subscale**.

3.1. Effect of DIF

Both the receptivity subscale and the mistrust subscale had items that functioned differentially. To examine the effect of the DIF in the two subscales, DIF-adjusted mean scores were compared across subgroups, as were the observed mean scores (Table 3).

In the receptivity subscale, DIF-adjustment of the sum score did not change the conclusion that there was no significant difference between the mean scores of parents with no previous child case and parents with a previous case (unadjusted scores: $p = 0.723$; adjusted scores: $p = 0.130$). In the mistrust subscale, DIF-adjustment provided stronger evidence for a difference in mistrust scores for parents with higher education of any length compared to parents with secondary education or less. Thus, there was strong evidence that parents with higher education had a lower degree of mistrust (mean: 14.98) than parents with secondary education or less (mean: 12.70) ($p = 0.002$). Had the scores not been adjusted for DIF, the evidence would have been weak ($p = 0.024$).

Tables for DIF-adjusting the summed scores are provided in Appendix D (Table D.1-2).

3.2. Targeting and reliability

The targeting of the scales was poor with the exception of the expectancy subscale where it was considered acceptable. Here the test information target indices indicated that on average 61.5 % of the maximum information was achieved in the expectancy subscale (Table 4). For the three other subscales, the information achieved relative to the maximally obtainable ranged between 41.0 % and 54.1 % depending on the subgroup of parents. The item maps in Appendix E (Fig E.1-4) illustrates the poor targeting, as person parameter estimates for some parents were located where item coverage was poor and thus in areas with less information than desired.

Reliability, taking any LD of the four subscales into account exceeded 0.90 for two subscales (expectancy: 0.94; Mistrust: 0.91). For the receptivity subscale the weighted mean reliability across DIF groups was 0.84, while the reliability of the working relationship subscale was 0.86.

Table 3
Comparison of observed and DIF-adjusted receptivity and mistrust scores across subgroups.

DIF groups (n)	Observed scores			Adjusted scores			
	Mean	SE	*p*	Mean	SE	*p*	Bias
Receptivity			0.723			0.130	
No previous case (76)	23.22	0.50		23.22	0.50		0
Previous case (134)	22.99	0.42		22.23	0.43		0.77
Mistrust			0.024[a]			0.002[a]	
Medium or long-cycle higher education	14.91	0.47		14.91	0.47		0
Vocational or short-cycle higher education	15.59	0.52		15.08	0.51		0.52
Secondary education or less	13.57	0.64		12.70	0.64		0.87

[a] There were no significant mean groups between the two groups of parents with higher educations, p-value refer to the difference between the combined group of parents with higher education (unadjusted mean 15.22, SE:0.35; adjusted mean 14.98, SE: 0.35) and parents with secondary education or less.

4. Discussion

In this study, we translated and conducted an initial validity study of a client engagement measure specifically within Denmark's child and family welfare system. Using RM and GLLRMs, we examined the psychometric properties of subscales for five aspects of client engagement. With slight adaptions to Yatchmenoff's (Yatchmenoff, 2005) original pool of 37 items to fit a Danish Child and Family Welfare system context, construct or essentially construct valid subscales, that fitted either a pure RM (expectancy) or relatively simple GLLRMs (receptivity, mistrust and working relationship) where DIF and/or LD had been taken into account, resulted for four out of five aspects of client engagement. For the last aspect of client engagement - investment, it was, however, not possible to fit a scale to a GLLRM. Reliability of all four subscales was high to excellent; however, the targeting of the scales to the study sample was generally poor except for the expectancy subscale.

In order to reach fit to an RM or a GLLRM, it was necessary to eliminate items from all four subscales. Common to all but one (item 28) of the eliminated items was, that they were also eliminated by Yatchmenoff's (Yatchmenoff, 2005) to develop the 19-item CES from the original pool of 37 items. This indicates that these items do not reflect the intended constructs in the Danish context and maybe not in a North American context either. In the Danish context some items that were eliminated from Yatchmenoff's (Yatchmenoff, 2005) final receptivity subscale (items 15 and 27) and mistrust subscale (item 30) were, however, retained. Thus, the 6-item receptivity subscale fitted a simple GLLRM with two instances of DIF relative to having had a previous child case and the 4-item mistrust subscale fitted a GLLRM with one instance of DIF relative to education and two locally dependent items. One possible reason for these differences is that these items could be kept in the Danish subscales due to the use of GLLRM where DIF and LD can be taken into account. Another reason may be that changing the response category 'not sure' into a less ambiguous midpoint labeled 'neither agree nor disagree' improved the functioning of the items altogether. However, an alternative explanation may also be that items excluded from Yatchmenoff's (Yatchmenoff, 2005) final receptivity (item 15: '*I see the same problems in our family that [CWS]/my caseworker does*' and item 27: '*I don't want anything to do with [CWS]*') and mistrust subscales (item 30:' *I always feel like I have to watch what I say around people from the agency*') function in the Danish CWS context, where the dynamics of client engagement may be influenced by an earlier initiated and preventive approach that prioritizes closely collaboration with the parents and preserving families (Pösö et al., 2014).

While Yatchmenoff (Yatchmenoff, 2005) chose to combine expectancy and investment into a joint buy-in subscale, these two aspects of client engagement were examined as separate subscales in the present study. This choice was based on conceptual considerations of expectancy and investment being separate concepts (although closely related) and that a joint scale would violate the psychometric criterion of unidimensionality. The 5-item expectancy subscale fitted a pure RM after elimination of two items. Four out of these five items were likewise included in Yatchmenoff's final 8-item buy-in subscale. In the Danish context, none of the 10 items of the investment subscale made up a scale adhering to the RM or GLLRM requirements. A likely explanation is that some items were ambiguously phrased, making it unclear whether they reflect genuine investment in the helping process – such as active participation in services, goal ownership and active help seeking – or simply the burden of interacting with CWS. For instance, item 35, '*I put a lot of time and effort into working with CWS,*' could indicate a parent's commitment to engagement, but could just as well reflect frustration

Table 4
Targeting and reliability of the four subscales.

Subscale (subgroups defining DIF)	Theta									Score				
	Target	Mean	TI mean	TI max	TI target index	RMSE mean	RMSE min	RMSE target index	PSI	Target	Mean	Target SEM	Mean SEM	r
Expectancy	−1.27	210	1.773	2.882	0.615	0.786	0.589	0.750	1.073	12.48	17.48	1.70	1.29	0.94
Receptivity[a]														
No previous case	−0.76	76	3.414	6.525	0.523	0.579	0.391	0.676	0.804	14.92	23.22	2.55	1.81	0.83
Previous case	−0.76	134	3.614	7.194	0.502	0.568	0.373	0.656	0.819	14.42	22.99	2.68	1.85	0.85
Mistrust[b]														
Medium or long-cycle higher education	−0.58	98	1.614	2.985	0.541	0.895	0.579	0.647	1.111	10.71	14.91	1.73	1.19	0.92
Vocational or short-cycle higher education	−0.40	59	1.944	3.956	0.491	0.775	0.503	0.648	1.008	11.54	15.59	1.99	1.32	0.89
Secondary education or less	−0.95	53	2.138	4.048	0.528	0.744	0.497	0.668	0.986	10.73	13.57	2.01	1.40	0.90
Working relationship	−0.65	210	1.616	3.937	0.41	0.822	0.504	0.613	1.016	8.66	11.02	1.98	1.21	0.86

Note. Targeting and reliability is presented for DIF subgroups in case of DIF. Targeting and reliability for DIF groups and subgroups in variables associated with the latent variable is available on request. TI: test information; RMSE: root mean-squared error of the estimated theta score; SEM: standard error of the measurement of the observed score; r: reliability.

[a] Weighted mean PSI across subgroups = 0.83; weighted mean reliability = 0.84.

[b] Weighted mean PSI across subgroups = 0.84; weighted mean reliability = 0.91.

with the demands of the system. If so, this item would not serve as a reliable indicator of engagement. Similarly, items such as *"I'm only doing what I have to do to keep my kids (get my kids back). Otherwise, forget it."* and *"I'm just doing whatever it takes to get [CWS] out of my life."* likely resonate most with families experiencing significant conflict with social services. Thus, research finds that parents' negative experiences and emotions when interacting with social authorities are well-documented, including feelings of being undermined, invisible, and not taken seriously (Thrana & Fauske, 2014). Families with low socioeconomic status, in particular, report the most negative experiences in these encounters (Kojan & Storhaug, 2021). In such interactions, parents often experience fear, stress, and anxiety (Kojan & Storhaug, 2021; Schreiber et al., 2013) – amid an already strained life situation – where their emotional experiences are shaped by negative expectations of collaboration and fear related to the power social authorities hold over their lives (Havnen et al., 2020). In contrast, our sample appears to consist of parents who are relatively satisfied with their interactions with the child and welfare services. This may explain why these items did not function as expected in our analyses, as they may not adequately capture the investment dimension in a sample where families generally perceive their relationship with CWS more positively.

A closer examination of the mistrust scores revealed significant differences based on educational background. Thus, parents with a shorter education (secondary education or less) exhibited a higher degree of mistrust compared to those with longer education (vocational or higher education of varying length). This is in accordance with previous studies finding parent engagement in welfare services to be associated with socio-demographic factors but also cultural barriers and marginalization (Kemp et al., 2009). Lower educational attainment or low income are associated with a higher risk for experiencing stressors such as financial difficulties, mental health problems and substance abuse (Galea et al., 2004; Silva et al., 2016). These factors have been linked to a lower engagement with services and may explain a heightened mistrust toward authorities, potentially shaping how parents perceive and interact with child welfare services (Kemp et al., 2009)

4.1. Strengths and limitations

A key limitation to the study is the relatively small sample size, which may have limited our ability to detect problems that a larger validation study could identify due to better power. However, as we were able to detect many and serious problems with the items, we can only assume that increased power, would have divulged even more issues. To our knowledge, this is the first study to validate a measure of client engagement in the Danish CWS context and thus constitutes an initial validation.

Although Yatchmenoff (Yatchmenoff, 2005) developed and investigated the structural validity of the CES using confirmatory factor analysis, they did not investigate the quality of the items, i.e., in terms of DIF, invariance, homogeneity or LD between items. Thus, it is unknown whether and to which degree Yatchmenoff 's CES subscales are affected by DIF. If affected by DIF, simply adding the items of the subscales to a sum score may lead to biased measurement which may cause biased analyses. The absence of prior studies assessing potential measurement biases makes the Rasch-based approach in the present study a major strength. The use of GLLRMs is particularly valuable, as it not only allows for the detection of DIF and LD, but also allows for adjustments to take into account uniform DIF and LD, rather than simply discarding items. Therefore, the use of these models ensure that the maximum number of items are retained, thus preserving items that are content valid, and maximizing reliability through the higher number of content valid and well-functioning items.

The reliability of all four subscales for the Danish sample were high to excellent with the expectancy subscale and mistrust subscale exceeding the 0.90 that is usually recommended for screening tools. Comparatively only the buy-in subscale in the 19-item CES with the American sample reached the level of 0.90 Yatchmenoff's (Yatchmenoff, 2005). However, as Yatchmenoff (Yatchmenoff, 2005) did not examine for DIF, invariance nor LD, the Cronbach Alphas reported in her study may be overestimated. Although reliabilities of the subscales were high to excellent in our study, it should be remembered that targeting was poor for three subscales – the expectancy subscale being the exception. While there is (almost) enough variation for the scales to function as screening tools, the questions are not well suited for the Danish target

sample, as the items are 'too easy' to endorse. This issue is most pronounced in the working relationship subscale, which showed the poorest targeting. This calls for additional, more 'difficult' items, i.e. items that require a higher level of engagement from parents to be endorsed. Additionally, the mean scores of the parents included in the sample indicate a relatively high level of engagement (Table 4). Although caseworkers were asked to recruit both cooperative and less engaged parents, parents with more severe cases or lower engagement may have been less likely to participate when invited by the research team. Therefore, we recommend that future validity studies also include a wider range of parents where less engaged parents are also represented in order to examine the targeting in a broader sample.

4.2. Conclusion and future directions

This initial validation of client engagement measures resulted in four construct or essentially construct valid subscales showing that expectancy, receptivity, mistrust and working relationship can be assessed within Danish CWS. In combination with high to excellent reliabilities, this indicates that the expectancy subscale and also the receptivity, mistrust and working relationship subscales, when adjusted for DIF and/or LD, may be relevant tools for use in large-scale research studies and in the long term potentially contribute to development of effective social practices among children and families of risk. However, the relatively small sample size and poor targeting highlight the need for further validity studies and refinement in larger and more diverse samples not least before these subscales can be reliably implemented in everyday child and family welfare practice where a precise assessment at the individual level is of utmost importance for the collaboration between parents and caseworkers.

CRediT authorship contribution statement

Signe Boe Rayce: Writing – review & editing, Writing – original draft, Validation, Investigation, Formal analysis, Data curation, Conceptualization. **Anne Marie Villumsen:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Maiken Pontoppidan:** Writing – review & editing, Investigation, Conceptualization. **Tine Nielsen:** Writing – review & editing, Writing – original draft, Validation, Formal analysis, Conceptualization.

Interest statement

The authors declare that they have no conflicts of interest.

Ethics

The study was conducted in accordance with the ethical standards of the with the Helsinki Declaration and has been approved by the Danish Center for Social Science Research's Research Ethics Committee (Journal no. 2024/3). All participants received written information about the survey, including details on the voluntary nature of participation, strict confidentiality of their data, and compliance with GDPR for data storage and analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to express our heartfelt gratitude to the participating families and social workers, whose contributions were essential to the success of this work. The families played a crucial role by both testing the questionnaire and participating in the survey, while the social workers provided invaluable support by reaching out to families and obtaining their consent for us to send the survey. In addition, we would like to thank Line Berg, Associate Professor, VIA University College, for her significant contribution to recruiting municipalities and the translation and adaption process. We thank student assistant Amalie Bøg Rasmussen for help with administrating the survey and Serkan Korkmaz for help with data preparation. Finally, we thank TrygFonden for funding this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chipro.2025.100267.

References

Andersen, E. B. (1973). A goodness of fit test for the rasch model. *Psychometrika, 38*, 123–140.
Bekaert, S., Paavilainen, E., Schecke, H., Baldacchino, A., Jouet, E., Zabłocka–Żytka, L., Bachi, B., Bartoli, F., Carrà, G., & Cioni, R. M. (2021). Family members' perspectives of child protection services, a metasynthesis of the literature. *Children and Youth Services Review, 128*, Article 106094.
Benbenishty, R., Davidson-Arad, B., López, M., Devaney, J., Spratt, T., Koopmans, C., Knorth, E. J., Witteman, C. L. M., Del Valle, J. F., & Hayes, D. (2015). Decision making in child protection: An international comparative study on maltreatment substantiation, risk assessment and interventions recommendations, and the role of professionals' child welfare attitudes. *Child Abuse & Neglect, 49*, 63–75.
Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57*(1), 289–300.
Bouma, H., Grietens, H., López López, M., & Knorth, E. J. (2020). Learning from parents: A qualitative interview study on how parents experience their journey through the Dutch child protection system. *Child & Family Social Work, 25*, 116–125.
Brandon, M., Philip, G., & Clifton, J. (2019). Men as fathers in child protection. *Australian Social Work, 72*(4), 447–460.
Charest-Belzile, D., Drapeau, S., & Ivers, H. (2020). Parental engagement in child protection services: A multidimensional, longitudinal and interactive framework. *Children and Youth Services Review, 116*, Article 105162.
Cheng, T. C., & Lo, C. C. (2016). Linking worker-parent working alliance to parent progress in child welfare: A longitudinal analysis. *Children and Youth Services Review, 71*, 10–16.
Christensen, K. B. (2013). Estimation of item parameters. *Rasch Models in Health*, 49–62.
Coady, N., Hoy, S. L., & Cameron, G. (2013). Fathers' experiences with child welfare services. *Child & Family Social Work, 18*(3), 275–284.
Cox, D. R., Spjøtvoll, E., Johansen, S., van Zwet, W. R., Bithell, J. F., Barndorff-Nielsen, O., & Keuls, M. (1977). The role of significance tests [with discussion and reply]. *Scandinavian Journal of Statistics*, 49–70.
Damiani-Taraba, G., Dumbrill, G., Gladstone, J., Koster, A., Leslie, B., & Charles, M. (2017). The evolving relationship between casework skills, engagement, and positive case outcomes in child protection: A structural equation model. *Children and Youth Services Review, 79*, 456–462.
Estefan, L. F., Coulter, M. L., VandeWeerd, C. L., Armstrong, M., & Gorski, P. (2012). Receiving mandated therapeutic services: Experiences of parents involved in the child welfare system. *Children and Youth Services Review, 34*(12), 2353–2360.
Fernandez, E. (2004). Effective interventions to promote child and family wellness: A study of outcomes of intervention through Children's family centres. *Child & Family Social Work, 9*(1), 91–104.
Fischer, G. H., & Molenaar, I. W. (1995). *Rasch models. FOUNDATIONS, recent developments, and applications*. Springer-Verlag.
Galea, S., Nandi, A., & Vlahov, D. (2004). The social epidemiology of substance use. *Epidemiologic Reviews, 26*(1), 36–52.
Gilbert, N. (1997). *Combatting child abuse: International perspectives and trends*.
Gilbert, N., Parton, N., & Skivenes, M. (2011). *Child protection systems: International trends and orientations*. Oxford University Press.
Gladstone, J., Dumbrill, G., Leslie, B., Koster, A., Young, M., & Ismaila, A. (2012). Looking at engagement and outcome from the perspectives of child protection workers and parents. *Children and Youth Services Review, 34*(1), 112–118.
Gladstone, J., Dumbrill, G., Leslie, B., Koster, A., Young, M., & Ismaila, A. (2014). Understanding worker–parent engagement in child protection casework. *Children and Youth Services Review, 44*, 56–64.
Greve, B. (2007). What characterise the Nordic welfare state model. *Journal of Social Sciences, 3*(2), 43–51.
Hamon, A., & Mesbah, M. (2002). Questionnaire reliability under the rasch model. In *Statistical methods for quality of life studies: Design, measurements and analysis* (pp. 155–168). Springer.
Hanson, B. A. (1998). Uniform DIF and DIF defined by differences in item response functions. *Journal of Educational and Behavioral Statistics, 23*(3), 244–253.
Havnen, K. J. S., Christiansen, Ø., Ljones, E. H., Lauritzen, C., Paulsen, V., Jarlby, F., & Vis, S. A. (2020). *Å medvirke når barnevernet undersøker. En studie av barn og foreldres medvirkning i barnevernets undersøkelsesarbeid*.

Johnson, J. M., Bristow, D. N., & Schneider, K. C. (2004). Did you not understand the question or not? An investigation of negatively worded questions in survey research. *Journal of Applied Business Research, 20*(1).
Kelderman, H. (1984). Loglinear Rasch model tests. *Psychometrika, 49*(2), 223–245.
Kemp, S. P., Marcenko, M. O., Hoagwood, K., & Vesneski, W. (2009). Engaging parents in child welfare services: Bridging family needs and child welfare mandates. *Child Welfare*, 101–126.
Kemp, S. P., Marcenko, M. O., Lyons, S. J., & Kruzich, J. M. (2014). Strength-based practice and parental engagement in child welfare services: An empirical examination. *Children and Youth Services Review, 47*, 27–35.
Khoo, E. G., Hyvönen, U., & Nygren, L. (2003). Gatekeeping in child welfare: A comparative study of intake decisionmaking by social workers in Canada and Sweden. *Child Welfare*, 507–525.
Kojan, B. H., & Storhaug, A. S. (2021). Barnevern og sosioøkonomisk ulikhet–sammenhenger, forståelser og ansvar. *Rapportserie for Sosialt Arbeid, 6*, 1–149.
Kreiner, S. (2003). *Introduction to DIGRAM. Department of biostatistics*. Copenhagen Copenhagen: University of.
Kreiner, S. (2007). Validity and objectivity: Reflections on the role and nature of Rasch models. *Nordic Psychology, 59*(3), 268–298.
Kreiner, S. (2011). A note on item–restscore association in rasch models. *Applied Psychological Measurement, 35*(7), 557–561.
Kreiner, S. (2013). The rasch model for dichotomous items. *Rasch Models in Health*, 5–26.
Kreiner, S., & Christensen, K. B. (2002). Graphical rasch models. In *Statistical methods for quality of life studies: Design, measurements and analysis* (pp. 187–203). Springer.
Kreiner, S., & Christensen, K. B. (2004). Analysis of local dependence and multidimensionality in graphical loglinear rasch models. *Communications in Statistics - Theory and Methods, 33*(6), 1239–1276.
Kreiner, S., & Christensen, K. B. (2007). Validity and objectivity in health-related scales: Analysis by graphical loglinear rasch models. In *Multivariate and mixture distribution rasch models: Extensions and applications* (pp. 329–346). Springer.
Kreiner, S., & Christensen, K. B. (2013). Person parameter estimation and measurement in rasch models. In *Rasch models in health* (pp. 63–78). Wiley Online Library.
Kreiner, S., & Nielsen, T. (2013). *Item analysis in DIGRAM 3.04. Part I: Guided tours. Research report 2013/06*. Department of Public Health Copenhagen, Denmark: University of Copenhagen.
Kreiner, S., & Nielsen, T. (2023). *Item analysis in DIGRAM 5.01*. Guided Tours.
Križ, K., & Skivenes, M. (2013). Systemic differences in views on risk: A comparative case vignette study of risk assessment in England, Norway and the United States (california). *Children and Youth Services Review, 35*(11), 1862–1870.
Lee, C. D., & Ayón, C. (2004). Is the client-worker relationship associated with better outcomes in mandated child abuse cases? *Research on Social Work Practice, 14*(5), 351–357.
Littell, J. H. (2001). Client participation and outcomes of intensive family preservation services. *Social Work Research, 25*(2), 103–113.
Maiter, S., Palmer, S., & Manji, S. (2006). Strengthening social worker-client relationships in child protective services: Addressing power imbalances and 'ruptured'relationships. *Qualitative Social Work, 5*(2), 161–186.
Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174.
Merenda, P. F. (2005). Cross-cultural adaptation of educational and psychological testing. In K. Hambleton Ronald, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 321–341). Lawrence Erlbaum Associates Publishers.
Merkel-Holguin, L., Hollinshead, D. M., Hahn, A. E., Casillas, K. L., & Fluke, J. D. (2015). The influence of differential response and other factors on parent perceptions of child protection involvement. *Child Abuse & Neglect, 39*, 18–31.
Mesbah, M., & Kreiner, S. (2013). Rasch models for ordered polytomous items. In K. B. Christensen, S. Kreiner, & M. Mesbah (Eds.), *Rasch models in health* (pp. 27–42). John Wiley & Sons.
Mirick, R. G. (2014a). Engagement in child protective services: The role of substance abuse, intimate partner violence and race. *Child and Adolescent Social Work Journal, 31*, 267–279.
Mirick, R. G. (2014b). The relationship between reactance and engagement in a child welfare sample. *Child & Family Social Work, 19*(3), 333–342.
Moesby-Jensen, C. K., & Moesby-Jensen, T. (2019). Forældre til børn med autismespektrumforstyrrelser eller ADHD og deres oplevelser af samarbejdet med myndighedsrådgivere. In *Diagnoser i myndighedsarbejde: Børn og unge med autisme og ADHD* (pp. 177–204). Samfundslitteratur.
Nielsen, T., & Kreiner, S. (2013). Improving items that do not fit the Rasch model. *Rasch Models in Health*, 317–334.
Oakland, T. (2005). Selected ethical issues relevant to test adaptations. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 65–92). Lawrence Erlbaum Associates Publishers.
Ólafsson, S. (2013). Well-Being in the nordic countries: An international comparison. *Icelandic Review on Politics & Administration, 9*(2).
Pösö, T., Skivenes, M., & Hestbæk, A.-D. (2014). Child protection systems within the Danish, Finnish and Norwegian welfare states—time for a child centric approach? *European Journal of Social Work, 17*(4), 475–490.
Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (Vol. 1993). Copenhagen: Danish Institute for Educational Research.
Rosenbaum, P. R. (1989). Criterion-related construct validity. *Psychometrika, 54*(4), 625–633.
Samsonsen, V., & Willumsen, E. (2015). Narratives from parents in England and Norway: Power and emotions in child protection assessments. *Journal of Comparative Social Work, 10*(1), 6–30.
Schreiber, J. C., Fuller, T., & Paceley, M. S. (2013). Engagement in child protective services: Parent perceptions of worker skills. *Children and Youth Services Review, 35*(4), 707–715.
Silva, M., Loureiro, A., & Cardoso, G. (2016). Social determinants of mental health: A review of the evidence. *The European Journal of Psychiatry, 30*(4), 259–292.
Sireci, S. G. (2005). Using bilinguals to evaluate the comparability of different language versions of a test. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 117–138). Lawrence Erlbaum Associates Publishers.
Smithson, R., & Gibson, M. (2017). Less than human: A qualitative study into the experience of parents involved in the child protection system. *Child & Family Social Work, 22*(2), 565–574.
Studsrød, I., Willumsen, E., & Ellingsen, I. T. (2014). Parents' perceptions of contact with the N orwegian C hild W elfare S ervices. *Child & Family Social Work, 19*(3), 312–320.
Tembo, M. J., & Studsrød, I. (2019). Parents' emotional experiences of their contact with the child welfare services: A synthesis of previous research-a research review. *Nordic Social Work Research, 9*(2), 184–198.
Thrana, H. M., & Fauske, H. (2014). The emotional encounter with child welfare services: The importance of incorporating the emotional perspective in parents' encounters with child welfare workers. *European Journal of Social Work, 17*(2), 221–236.
Trotter, C. (2002). Worker skill and client outcome in child protection. *Child Abuse Review: Journal of the British Association for the Study and Prevention of Child Abuse and Neglect, 11*(1), 38–50.
Xu, Y., Ahn, H., & Bright, C. L. (2017). Family involvement meetings: Engagement, facilitation, and child and family goals. *Children and Youth Services Review, 79*, 37–43.
Yatchmenoff, D. K. (2005). Measuring client engagement from the client's perspective in nonvoluntary child protective services. *Research on Social Work Practice, 15*(2), 84–96.