



# TESTING THE RELIABILITY AND VALIDITY OF THE GRADED CARE PROFILE VERSION 2 (GCP2)

Robyn Johnson, Emma Smith and Helen Fisher  
NSPCC Evaluation department

October 2015

**NSPCC**

EVERY CHILDHOOD IS WORTH FIGHTING FOR

# Impact and Evidence series

This report is part of the NSPCC's Impact and Evidence series, which presents the findings of the Society's research into its services and interventions. Many of the reports are produced by the NSPCC's Evaluation department, but some are written by other organisations commissioned by the Society to carry out research on its behalf. The aim of the series is to contribute to the evidence base of what works in preventing cruelty to children and in reducing the harm it causes when abuse does happen.

# CONTENTS

ACKNOWLEDGEMENTS	4
KEY FINDINGS	5
KEY FINDINGS: YOUNG PEOPLE'S VERSION	6
EXECUTIVE SUMMARY	8
MAIN REPORT	12
Introduction	12
Context	12
Methodology	16
Results	22
Discussion	34
References	40
APPENDICES	42
Appendix A: Evidence for use of other tools	42
Appendix B: GCP2/Ncfas-G/Home Inventory mapping	43
Appendix C: Additional Tables	45
Appendix D: Face validity	49

# ACKNOWLEDGEMENTS

The authors would like to thank:

- The families who agreed for their information to be used for the testing.
- The NSPCC practitioners and managers who took part in the testing and supported the evaluation in a number of ways such as seeking consent from families.
- The Commission Operational Group for their advice, support and guidance in relation to this evaluation, especially Dawn Hodson, Claire Gorman and John Harrison.
- Dr Srivastava and Dr Richard Fountain for their support of the testing.
- Richard Cotmore for his advice and support, particularly with the report.

# KEY FINDINGS

- The second version of the Graded Care Profile (GCP2) was tested for reliability and validity.
- Inter-rater reliability was assessed by comparing GCP2 scores between pairs of practitioners. There was a substantial level of agreement for the domains of Physical Care and Developmental Care, and a moderate level of agreement for Care of Safety and Emotional Care.
- Concurrent validity was assessed by comparing GCP2 scores to scores obtained using two other validated measures. There were strong correlations between the GCP2 and the two other tools across all domains.
- Face validity was measured by seeking the views of professionals with regards to the extent to which the GCP measures what it intends to measure. The results suggest that the GCP has a fairly high level of face validity, with participants generally agreeing that the GCP achieves what it sets out to do.
- In conclusion, GCP2 has been found to be reliable and valid. It can be used in the knowledge that it has sound psychometric properties, and is a reliable and valid assessment tool in aiding practitioners in the assessment of child neglect.

# KEY FINDINGS: YOUNG PEOPLE'S VERSION

Graded Care Profile (GCP) is a way of helping social workers understand how well a child is being cared for. GCP helps social workers spot things that might be putting a child at risk of harm by giving them useful questions to ask parents and families. This means that it is easier for the child to get the help they need.

GCP is an example of a tool that social workers use. All tools should be “tested” to make sure they do what they are supposed to do. The NSPCC had collected people’s comments about the GCP and how it could be improved. Some changes were then made to the GCP and it is now called ‘GCP2’. The NSPCC tested GCP2 to make sure that it is a good tool. To do this they tested:

- Its **reliability** to see if the results from the GCP2 are the same when it is used by two different people. To do this, they asked two different social workers to use GCP2 with the same families.
- Its **validity** to find out if it actually measures what it’s supposed to measure. To do this, they asked social workers to fill in the GCP2, as well as another tool for the same family. They also asked some social workers if they thought GCP2 does what it’s supposed to do.

The NSPCC found that:

- GCP2 is a **reliable** tool, because the two different social workers usually had the same results when they each used the GCP2 with the same family.

- GCP2 is a **valid** tool, because the results were similar when the same social worker used the GCP2 and another tool with the same family. Social workers also told the NSPCC that they thought the GCP2 generally does what it's supposed to do.

Social workers and other professionals can use GCP2 knowing that it has been tested and has been found to be a good tool in helping them understand how well a child is being cared for.

# EXECUTIVE SUMMARY

## Context

The Graded Care Profile (GCP) scale was developed in 1995 as a practical tool to give an objective measure of the care of a children across all areas of need where there are concerns about neglect. Care is graded across four domains: A) Physical Care, B) Care of Safety, C) Emotional Care, and D) Developmental Care. Some psychometric testing was carried out in 1997, which found high levels of inter-rater reliability, although there were some gaps in the testing. There is evidence pointing to the GCP's potential and the ways in which it helps practitioners' practice in assessing neglect, including evidence from the NSPCC's evaluation of the tool (Johnson and Cotmore 2015). However, some limitations of the tool were also identified, and improvements suggested, largely in relation to the language used in the tool. A review group was formed, comprising Dr Srivastava, the GCP developer, his colleague Richard Fountain and representatives from the NSPCC. The group amended the GCP in various ways, although the core principles remained the same. This resulted in the second version of the tool, which is now known as 'GCP2'. The aim of this piece of work was to test the reliability and validity of GCP2.

## Approach

The testing was carried out through NSPCC service centres, with three pairs of social work qualified practitioners. Testing took place in a service setting, and as such there were tensions between the requirements of social work practice and those of robust testing. These tensions were addressed and

agreements reached in ways that were comfortable for the practitioners in their practice but which did not undermine the robustness of the testing process. Examples of this are discussed in the full report.

### Reliability testing:

**Inter-rater reliability** was assessed by comparing GCP2 scores assigned by each practitioner in a pair, when used in relation to the same child. Inter-rater reliability was calculated using the Kappa statistic to examine the level of agreement between raters. The sample consisted of 30 children, with an average age of seven. All children were classified as either ‘Child in Need’ or ‘Child Protection procedures or plans’.

### Validity testing:

**Concurrent validity:** Practitioners scored the same child on both the GCP2 and two other validated measures: North Carolina Family Assessment Scale (General version) (NCFAS-G) and the Home Observation for the Measurement of the Environment Inventory (known as the HOME Inventory). There is a close match between these measures and the domains in the GCP2. GCP2 scores were compared to corresponding scores in both NCFAS-G and the Home Inventory. Spearman’s rho correlation was used to assess the extent to which the scores correlated. There were 15 children in the sample, with an average age of nine. All children were classified as either ‘Child in Need’ or ‘Child Protection procedures or plans’.

**Face validity:** This was measured by obtaining the views of nine professionals in relation to the extent to which the GCP measures what it intends to measure.

## Results

**Reliability testing:** There was a substantial level of agreement for the Physical Care domain ( $k=0.65$ , Confidence Interval (CI) =0.43 to 0.87), a moderate level of agreement for Safety ( $k= 0.56$ , CI= 0.36 to 0.77), moderate agreement for Emotional Care ( $k= 0.57$ , CI= 0.50 to 0.89) and substantial agreement for Developmental care ( $k=0.69$ , CI= 0.50 to 0.89). A breakdown of results for each pair of practitioners is provided in the full report.

## Validity testing

**Concurrent validity:** There were strong correlations between the GCP2 and the two other tools across all domains. The correlations between the GCP2 areas and the main related areas from the other tools were strong and statistically significant and ranged from 0.55 to 0.91. The vast majority of the correlations of the related sub-areas were also statistically significant.

**Face validity:** The GCP2 can be said to have a fairly high level of face validity, with participants generally agreeing that the GCP2 achieves what it sets out to do. However, the results suggest that it may be beneficial to review the extent to which the GCP2 aims to enable the assessment of child development and developmental disorder, as well as educational, emotional and behavioural outcomes. The evidence also suggests that it would be helpful to consider perceived gaps, particularly how the impact of parental problems on children are reflected in the scoring.

## Conclusion

GCP2 has been found to be reliable and valid. It can be used in the knowledge that it has sound psychometric properties, and is a reliable and valid assessment tool in aiding practitioners in the assessment of child neglect.

# MAIN REPORT

## Introduction

The second version of the Graded Care Profile (GCP) has been tested for reliability and validity. This report describes why it was tested, how the testing was carried out and reports on the findings.

## Context

The GCP was designed by Dr Srivastava, a paediatrician from Luton, in 1995 as a practical tool to give an objective measure of the care of children across all areas of need where there were concerns about neglect. This tool has now been deployed in over 60 local authorities to help identify and intervene in cases of child neglect. The tool draws on the concept of a continuum of care rather than compartmentalising care into neglectful and non-neglectful. There are five grades, which are based on three different factors: the level of care, commitment to care, and the quality of care. The grades are applied to Maslow's model of human needs: physical, safety, love and belongingness, and esteem (Srivastava and Polnay 1997). There are four domains (areas) in the tool. These were originally: A) Physical care, B) Safety, C) Esteem, and D) Love. Following amendment of the tool in 2015 (see below), the domains were renamed slightly and are now known as A) Physical Care, B) Care of Safety, C) Emotional Care, and D) Developmental Care.

The originators of the tool carried out inter-rater reliability testing soon after it was developed in 1997. Nursery children were scored independently

by two different raters. They found an almost perfect level of agreement in the areas of physical care, safety and esteem, and a substantial level of agreement in love, and concluded that “This scale appeared user-friendly and provided grading of care with high inter-rater agreement.” (Srivastava and Polnay, 1997). However, the testing only involved nursery-aged children, so it has not been tested with other age groups. It also included only health visitors and nursery workers, so Children’s Social Services were not included – and this is the arena it appears to be most commonly used in. In addition, validity testing was not carried out. There were, therefore, gaps in the original psychometric testing of the measure, and further testing was required.

In 2012, Barlow et al carried out a review of structured professional judgment, which found GCP to be one of only two “promising” assessment approaches, as judged by the evidence of their testing and development. Nonetheless, they pointed to the need for further testing of it in terms of validity and reliability. Referring to GCP and another tool (Safeguarding Assessment and Analysis Framework), they said:

“[the tools] currently have limited evidence available concerning their rigour, and their use should be piloted and assessed for their validity, reliability and impact, in addition to the collection of data about service user/provider acceptability, and equity.”

Barlow et al. 2012, p10

There have been some positive findings in relation to the GCP, including from the NSPCC's evaluation of the tool (Johnson and Cotmore, 2015). The tool was generally seen by practitioners and managers as a useful tool for assessing the child's needs, and respondents said that it often played an important part in informing their case planning. This could result in professionals' concerns being alleviated or alternatively escalated. The GCP affected practice in other ways too, by: increasing practitioner skills; improving recording and reporting practices; and facilitating communication with a range of stakeholders, including with parents, managers and other professionals.

Professionals were able to identify several features of the GCP which they felt helped them to identify risks and potential harm more effectively. This included the scoring process, which they felt helped to specify the nature of the neglect and to quantify it. It could also serve to make the neglect more "visible" – to them, but also to parents and others. The child-centred nature of the GCP process also helped in the assessment of care, as practitioners were encouraged to focus on the child's experience which helped to disentangle them from the adults' agendas.

Evidence from the NSPCC evaluations suggested that the GCP assessment process could act as a change enabler. Aspects of the GCP which were seen as effective change enablers were: the scoring process, as it helped the parents understand the professionals' concerns; the child-centred focus, as it promoted the practitioner's understanding of the child's needs; a shared perception between the

professional and parent that the GCP was generating an objective judgment rather than the professional reaching a subjective one, resulting in parents feeling less criticised; and that professionals were enabled to communicate more clearly about the needs and change required. Hence the change enabling appears to happen through two channels:

- Improved understanding leading to a better informed plan;
- and a more constructive process for addressing the challenges resulting in improved motivation.

However, some limitations of the tool were also identified, and improvements suggested, largely in relation to the language used in the tool.

Practitioners' perceptions of the tool were collated and analysed by the evaluation team and a report prepared for the GCP review group. The review group comprised Dr Srivastava, the GCP developer, his colleague Richard Fountain and representatives from the NSPCC. Following a review of a range of evidence, including training feedback and other research as well as the evaluation findings, the GCP review group amended the GCP in various ways, although the core principles remained the same.

This resulted in the second version of the tool, which is now known as 'GCP2'. Amendments focused on the language used in the tool, gaps in the assessment of neglect, confirmation of scaling and scoring processes, and formatting issues. Due to these changes, the amended version of the GCP required testing. The second stage of the evaluation involved testing the reliability and validity of GCP2.

The aim of this stage of the evaluation is to assess the reliability and validity of the second edition of the GCP. The research questions are:

- How reliable is the second edition of the GCP?
- How valid is the second edition of the GCP?

## Methodology

### Recruitment

All cases were recruited through NSPCC service centres in England, which received referrals from their respective Local Authorities. Three pairs of practitioners took part in the testing, from three different service centres. All six practitioners involved had prior experience of using the GCP2, with at least two families. Practitioners sought consent from parents for their scores to be used for the testing, after receiving training and guidance. All referrals where there was a concern about neglect were accepted. Testing took place between September 2014 and August 2015.

A case was defined as each GCP2 carried out. Therefore, there could be more than one “case” per family, if the GCP2 was carried out with more than one child in a family. Practitioners were advised in their training to complete GCP2s for each child where there was a concern in relation to neglect. In addition, GCP2s could be completed with more than one parent in a family. Practitioners were asked to submit all GCP2s carried out.<sup>1</sup> A maximum of two GCP2s per family were accepted for analysis, to

---

1 Only one team submitted more than two GCP2s per family. For this site, cases were selected based on:

- Which child had been selected for validity testing
- Aiming to have one GCP2 from each parent of a different child
- Obtaining a representative mix of child age groups

enhance robust testing through the use of as many independent cases as possible. Scores were checked to ensure that differences between different family members were large enough to be considered as reasonably distinct cases, i.e. that the care of two siblings or by two parents was not identical. The threshold was at least a one point difference across at least two items by at least one rater.

## Reliability

The tool's reliability was assessed by carrying out inter-rater reliability testing, which has been defined as "*a measure of reliability used to assess the degree to which different judges or raters agree in their assessment decisions*" (Phelan and Wren 2006). In order to carry out this testing, the GCP2 was used by pairs of practitioners, who collected the evidence together but carried out the scoring separately. The same practitioners were involved throughout the testing so that it can be ascertained that it is the tool that determines how similar the ratings are rather than the individuals involved. Three pairs of practitioners (across three teams) used the GCP2 with 10 cases (the minimum required for analysis), so there is a total of 30 cases.

Practitioners scored as soon as they had enough information, for example, after each family visit, or when they agreed that they had enough information to score a whole domain. This approach was used in order to overcome an issue raised by practitioners, in that they needed to reflect on a case and discuss their visits during the course of the assessment, in line with their good practice guidelines. However, to fulfil the requirements of inter-rater reliability testing, discussion before scoring was avoided so that scoring was not unduly influenced by the other

practitioner in the pair. Once scoring for a domain/s had been done, practitioners were free to discuss the scoring for that domain/s, but no changes were made to scores.<sup>2</sup> The approach used was therefore a balance in attempting to meet the needs of both the reliability testing and good social work practice.

It was acknowledged that there may be occasions where it was unavoidable that one practitioner may influence the other, for example through their reactions to what is seen or heard in the family home, or due to instances where they may be required to intervene immediately for the safety of a child. Therefore, both practitioners were asked about the extent to which they thought their scores were influenced by their co-rater. For all cases included in this analysis, practitioners responded 'not at all'.

Practitioners were advised that both practitioners in a pair should endeavour to attend all visits together, but where this was not practically possible the practitioner who undertook the visit without the co-rater should describe what took place during the visit (what they observed/heard from family members) but they should not discuss their interpretation of the evidence. They were also advised that in carrying out a joint visit, both practitioners should endeavour to stay together during the visit where possible so that they were exposed to the same evidence within the family environment. As above, where this was not possible in practice, one should describe

---

2 Once all scores had been submitted for testing purposes, reflective case discussions took place in which moderated GCP2 scores were agreed, in conjunction with the practitioners' line manager. This was the version held on the case file and used for all work with the family. The original unmoderated GCP2 scores were not held on the case file and were classed as 'research data'.

the evidence to the other but not discuss their interpretation of it.

GCP2 scores submitted by pairs of practitioners were compared. The ratings were compared using Cohen's Kappa to assess inter-rater reliability. This statistic was used because the data are on an ordinal (categorical) scale rather than a continuous scale. Kappa scores are conventionally interpreted on the basis of one of six levels of agreement (Landis and Koch, 1977), as follows:

- < 0.00 Poor
- 0.00-0.20 Slight
- 0.21-0.40 Fair
- 0.41-0.60 Moderate
- 0.61-0.80 Substantial
- 0.81-1.00 Almost Perfect

The 'internal consistency'<sup>3</sup> of each domain was also calculated using Cronbach's alpha to see how well the items that should go together in a scale do actually hang together.

## Validity

Validity was assessed by testing concurrent validity and face validity.

**Concurrent validity** is demonstrated when an assessment tool correlates well with another tool measuring a similar construct that has previously been validated. This therefore involved testing the GCP2 alongside another assessment tool and then

---

3 Internal consistency: The degree to which each item measures the same construct. For example, a statistical test measured the degree to which each item making up 'Area A' (physical care) on the GCP2 related to the other item(s) measuring this construct in the GCP2.

comparing the scores between the two tools to see how similar they were. A combination of two tools were used because there is not another neglect assessment tool that covers all domains within the GCP2. The tools were:

- North Carolina Family Assessment Scale (General version) (NCFAS-G) (Kirk and Ashcraft, 1998). This tool is designed to examine family functioning in the domains of Environment, Parental Capabilities, Family Interactions, Family Safety, and Child Well-being. The whole of NCFAS-G was used, because permission was granted on the basis of the tool being used in its entirety.<sup>4</sup>
- The Home Observation for the Measurement of the Environment Inventory (known as the Home Inventory) (Caldwell and Bradley, 1984, 2003). The Home Inventory is an observer-rated measure designed to evaluate the quality and quantity of stimulation and support available to a child in the home environment. The focus is on the child in the environment, the child as a recipient of inputs from objects, events, and transactions occurring in connection with the family surroundings. Permission was obtained to use the two most relevant sections of this tool.<sup>5</sup>

These tools were selected because they have been shown to be reliable and valid (please see Appendix A), and between them, they most closely match the domains in the GCP2. A mapping exercise was carried out, whereby the sub domains and items of each tool were mapped conceptually

---

4 Thank you to Priscilla Martens, Executive Director, National Family Preservation Network

5 Thank you to Dr Robert Bradley at T. Denny Sanford School of Social and Family Dynamics.

based on the content of the tools (See Appendix B). This determined which items were compared during analysis.

One of the practitioners in the pair was assigned to the validity testing. This practitioner completed the other tools, in addition to the GCP2. The practitioner was asked to complete the NCFAS-G and Home Inventory in relation to the child that there was most concerns about.

GCP2 scores were compared to corresponding scores in both NCFAS-G and the Home Inventory, as identified during the mapping exercise.

Spearman's rho correlation was used to assess the extent to which the scores were correlated. This was used because it is not affected by differing scales. The relevant total scores and subscales were compared.

**Face Validity** ascertains that a measure appears to be assessing the intended construct under study, i.e. whether the test appears (at face value) to measure what it claims to. For GCP2, this meant testing whether the tool appears to assess neglect and quality of care. This was tested by obtaining feedback from practitioners, their line managers and an expert in neglect, via a questionnaire. Questions focused on the extent to which they felt the GCP2 fulfils its purpose, as set out in the GCP guidance (please see Table 4 below for questions asked). Quantitative data was analysed in Snap Professional, and open-ended comments were analysed thematically.

## Ethics

This evaluation was approved by the NSPCC Research Ethics Committee.

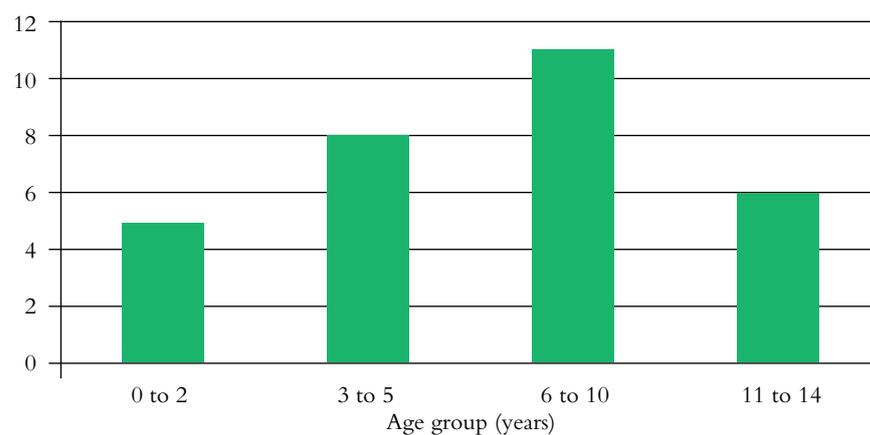
## Results

There was an average of six visits to each family by practitioners ranging from three to 11, including a few unannounced visits in some cases. Evidence was therefore collected over a number of visits across several weeks. All visits were attended by both practitioners in two sites; and 90% of visits in the third site.

## Reliability testing

For each site, the GCP2 was carried out with five families so a total of 15 families. In each family, GCP2 scores were analysed in relation to two children, so there were a total of 30 children in the sample. There was a fairly even gender split, with 47% female (N=14) and 53% male (N=16) children. There was a good spread of children's ages, ranging from infant to 14, as shown in figure 1. The mean age was 7 (standard deviation 4.03). All except two children were White British.

Figure 1. Distribution of children included in reliability testing by age group (N=30)



40% of cases were classified as 'Child in Need' and 60% as 'Child Protection procedures or plans'. Please see Appendix C (Table 6) for a breakdown

by site. There were no children where there were ‘potential concerns’ (i.e. cases where there were no pre-existing and confirmed concerns about neglect). However, across sub domains, 32% of scores in all cases were scores of 1 or 2, so there was a range of level of concerns, as the sample included families with very good care in at least some domains.

Inter-rater reliability was calculated using the Kappa statistic to examine the level of agreement between raters. The results for each area are illustrated in Table 1. Agreement appeared strongest for the areas of Physical Care and Developmental Care, as the levels were ‘substantial’ for the teams as a whole and between ‘moderate’ and ‘almost perfect’ for the teams individually. Overall agreement for both Care of Safety and Emotional Care was moderate, although Team A had only slight agreement for Care of Safety. Potential reasons for this difference are detailed in the discussion.

Considering each team separately, there were three Kappa scores that did not reach statistical significance: Team A ratings of Safety and Team B ratings of Physical Care and Emotional Care. However, agreement was statistically significant for all teams collectively across all four domains.

Internal consistency (measured by Cronbach’s Alpha) was excellent for Developmental Care, good for Physical Care and Emotional Care and acceptable for Care of Safety<sup>6</sup>.

---

6 Interpretation of the Cronbach’s alpha coefficients were based on the guidelines described by George and Mallery (2003): Above 0.9 = Excellent, Above 0.8 = Good, Above 0.7 = Acceptable, Above, 0.6 = Questionable, Above 0.5 = Poor, and less than 0.5 = Unacceptable.

Table 1: Agreement between raters for different GCP Areas

	<b>Team A Kappa (95% CI)</b>	<b>Team B Kappa (95% CI)</b>	<b>Team C Kappa (95% CI)</b>	<b>All Teams Kappa (95% CI)</b>	<b>Internal consistency: Cronbach's Alpha</b>
<b>A: Physical Care</b>	0.50* (0.21, 0.79)	0.40, NS (0.16, 0.96)	90% agreement <sup>a</sup>	0.65** (0.43, 0.87)	0.84
<b>B: Care of Safety</b>	0.14, NS (-0.30, 0.58)	90% agreement <sup>a</sup>	0.44* (-0.13, 1.01)	0.56** (0.36, 0.77)	0.73
<b>C: Emotional Care</b>	0.44* (0.06, 0.81)	0.39, NS (-0.07, 0.83)	0.81** (0.46, 1.16)	0.57** (0.50, 0.89)	0.88
<b>D: Developmental Care</b>	0.47** (0.13, 0.80)	0.67** (0.28, 1.05)	0.82** (0.52, 1.12)	0.69** (0.50, 0.89)	0.93

\* Significant at the 0.05 level. \*\*Significant at the 0.01 level. CI = Confidence intervals<sup>7</sup>, GCP2 = Graded Care Profile 2nd version, NS = Non-significant.

<sup>a</sup>There was 100% agreement in 90% of cases, so 'almost perfect'. Unfortunately there was not enough variance to calculate the Kappa score for this Area.

A breakdown of agreement for each sub area for all teams combined is provided in Table 7 in Appendix C.

Table 2 illustrates the proportion of scores where there was complete agreement between raters for each area. Area A had the highest percentage of complete agreements, followed by Area D, B and lastly Area C.

<sup>7</sup> Confidence intervals convey how precise the kappa value obtained is and are calculated based on the standard error. The narrower the confidence intervals are the more likely that the kappa value obtained is a good estimate of the agreement that would be found between any two practitioners from the wider population. The 95% confidence intervals presented here indicate that was agreement between 100 pairs of practitioners compared, 95 times the kappa value obtained would fall between these confidence intervals and five times the kappa value would be either higher or lower than the range of these confidence intervals.

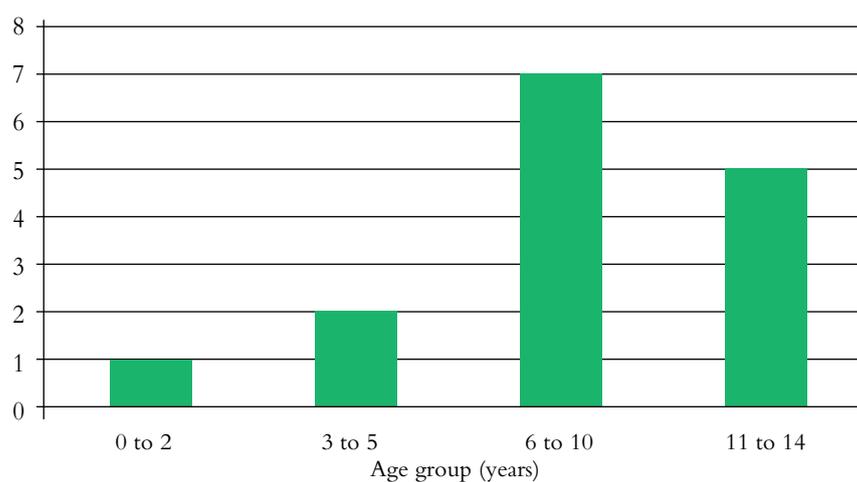
Table 2: Complete agreement in scores between the two raters

GCP2 Area	Complete agreement (%)
A	80
B	70
C	66.6
D	76.6

## Validity testing

Validity testing was carried out in relation to one child from each family (the child there was the most concerns about). Therefore, 15 children from 15 different families (5 per team) were used for the validity testing. Six children were girls and the vast majority (14/15) were White British. Only three children were aged 5 or under. The average age was 9 (standard deviation 3.5). The number of children in each age category is shown in Figure 2 below. Six out of the 15 children were classed as ‘Child in Need’ and nine as ‘Child Protection’ cases.

Figure 2. Distribution of children included in validity testing by age group (N=15)



Overall there were strong correlations between the GCP2 and the two other tools across all domains. The correlations between the main GCP2 areas and the main related areas from the other tools were strong and statistically significant and ranged from 0.55 to 0.91. All of the correlations of the related sub-areas were also statistically significant, apart from four items in GCP2 Area A (Physical Care) when compared to NCFAS-G (See Table 9 in Appendix C). However, Area A correlated strongly and significantly with the NCFAS-G Absence/Presence of neglect item (See Table 3). Therefore, the results suggest overall that the GCP2 is measuring similar concepts to those identified via the mapping exercise from NCFAS-G and the Home Inventory records.

Table 3: Correlations between the GCP2 and related NCFAS-G and HOME main areas

NCFAS-G/Home main areas	GCP2 main areas			
	Area A	Area B	Area C	Area D
Absence/presence of neglect (NCFAS-G)	- 0.77**	- 0.91**	- 0.55*	- 0.64*
A total (HOME)	-	- 0.55*	- 0.82**	- 0.79**
B Total (HOME)	-	- 0.67**	- 0.76**	- 0.72**

\*Spearman correlation coefficient is significant at the 0.05 level, \*\*Spearman correlation coefficient is significant at the 0.01 level. GCP2 = Graded Care Profile 2nd version, NCFAS-G = North Carolina Family Assessment Scale for General Services, HOME = Home Inventory Record Sheets.

## Face Validity

Face validity questionnaires were completed by a total of nine participants, comprised of six practitioners, two managers and one expert in the field of child neglect. The neglect expert was an experienced researcher with a long track record of research in the area of child protection and neglect who has published widely.

The results from these are presented in Table 4. Participants were asked to explain their responses, and these are discussed below the table.

Table 4: Face validity results – frequency of responses to each question.

	A large extent	To some extent	Not much	Not at all	No response
To what extent do you think GCP2 addresses <b>all areas of a child's needs</b> ?	4	5	0	0	0
To what extent do you think GCP2 enables the <b>level of neglect</b> to be measured?					
Physical care	7	2	0	0	0
Care of Safety	4	5	0	0	0
Emotional care	5	4	0	0	0
Developmental care	6	3	0	0	0
To what extent do you think GCP2 enables professionals to give an <b>objective</b> measure of a child's care?	5	4	0	0	0
To what extent do you think GCP2 enables professionals to accurately measure the <b>quality of care</b> ?	7	2	0	0	0
To what extent do you think GCP2 enables professionals to see things from the <b>child's perspective</b> (i.e. understand things from the child's viewpoint/what it's like for the child in the current home environment)?	4	3	1	0	1
Where neglect is suspected, to what extent do you think GCP2 enables <b>progress to be monitored</b> after interventions?	4	3	0	0	2
To what extent do you think GCP2 enables:					
<b>prevention</b> (early recognition and intervention)	2	5	0	1	1
<b>limitation</b> (timely referral and intervention)	6	2	0	0	1
<b>urgent protection</b>	3	4	1	0	1
To what extent are the <b>grading constructs</b> (i.e. the descriptions of each grade in the coding manual) accurately defined?	4	5	0	0	0
Where neglect is suspected, to what extent do you think GCP2 enables <b>intervention to be targeted</b> ?	7	1	1	0	0
Where quality of care is of interest, to what extent do you think GCP2:					
assesses a child's <b>development</b>	1	5	3	0	0
assesses <b>developmental disorder</b> in a child	0	5	3	1	0
enables a professional to understand <b>educational outcomes</b> in a child	0	6	2	1	0
enables a professional to understand <b>emotional or behavioural outcomes</b> in a child	2	6	0	1	0
Overall to what extent do you think GCP2 <b>measures what it intends to measure</b> ? (i.e. the quality of care by an individual child delivered by an individual carer or parent)	9	0	0	0	0
As far as you are aware, to what extent is GCP2 consistent with <b>current knowledge and best practices</b> in child neglect?	8	1	0	0	0

The results were generally positive with participants indicating that the GCP2 fulfils its purpose either to a large extent or to some extent, across the vast majority of its aims, apart from a specific set of aims as discussed below. Crucially, all participants agreed that the GCP2 measures what it intends to measure to a large extent.

All nine participants felt that the GCP2 addresses all areas of a child's needs, at least to some extent. Some referred to perceived "gaps" in the tool with one explaining that:

"the GCP is a neglect assessment, but neglect does not address all areas of a child's needs as it does not assess mental health, domestic abuse, substance misuse and the impact of these. A child's needs are not necessarily all to do with neglect e.g. sexual abuse,"

Practitioner

Additional perceived gaps are listed in Appendix D1. However, others explained that "if used properly" the GCP2 does address all needs, and referred to the skills of the practitioner involved:

"I believe with familiarity, experience and confidence a practitioner can incorporate almost any issue and ensure that this is 'encapsulated' within the body of the accompanying assessment report. areas missing i.e. impact of domestic abuse however this can be captured within the

safety in presence, absence emotional areas in my opinion.”

### Practitioner

Respondents considered that GCP2 enables the level of neglect to be measured, at least to some extent, across all four domains. However, there was a reminder that GCP2 is a “snapshot” during the assessment period and respondents highlighted the importance of including reasons for scoring in order to provide a context.

All participants felt that GCP2 enables professionals to give an objective measure of a child’s care, at least to some extent. The value of the grading constructs was highlighted here, as well as the fact that the scores are supported by observations and evidence. Those who felt that it was less objective referred to:

- the possibility of identifying ‘false positives’
- the wording in the grading constructs, e.g. the suggestions that some of the wording (especially for grade 5) are too prescriptive and so may not be used correctly (see Appendix D2)
- the potential impact of other influences, such as the information/ opinions of other professionals and level and nature of engagement of families

Participants were very positive about the extent to which GCP2 enables professionals to accurately measure the quality of care. They explained that GCP2 is “holistic”, considering both emotional and physical well-being, with the numerical grading facilitating the measurement of care. A manager commented:

“We understand that neglect is omission rather than commission. However, the tool helps explain what the child needs in order to reach full potential. It helps parents understand what they need to do in order to achieve ‘good parenting’. It helps professional support workers know what areas to focus support upon and how much help to provide.”

Manager

Interestingly there was a mix of perspectives with regards to the extent to which GCP2 enables professionals to see things from the child’s perspective. The uncertainty here centred on the following:

- Ensuring that adequate information is gleaned from the child:

“I still feel there is a gap where the child’s voice can be heard. On a number of our cases we have only been able to see the child on a few occasions. Therefore, I feel uncomfortable at times as to whether we have measured this accurately enough”

Practitioner

- Possible gaps due to the impact of parental problems (drugs/alcohol, domestic abuse) not being overtly addressed in the tool.

Interestingly, one practitioner reflected on the weight the child’s perspective should hold in the assessment:

“The grading points are very focussed on the child’s experience and what life is like for them. I think that the constructs are able to score what the child experiences are in terms of the level of care that they receive from their parents. However I don’t always think that’s necessarily what their view would be. For example I might score a 4 for housing décor because of dirtiness, clutter etc. but the child might not recognise this if they have always lived in that situation so their ‘point of view’ might be different from what I score using the grading constructs.”

#### Practitioner

There was a general consensus that the GCP2 enables progress to be monitored after interventions<sup>8</sup>, with participants explaining that it provides a benchmark, and enables progress or lack of it to be clearly demonstrated.

Participants felt that GCP2 enables appropriate action to be taken, whether that is prevention<sup>9</sup> (early recognition and intervention), limitation (timely referral and intervention), or urgent protection. There was a suggestion that the earlier the tool is used the better, to ensure timely improvements in the care of children (hence slightly less positive responses in relation to ‘urgent protection’). With regards to prevention, participants explained that the GCP2 helps to identify concerns, secure appropriate

---

8 There was some missing data here where participants had not had experience of using GCP2 at Time 2.

9 A small number were unsure about their response here, as they had only used GCP2 with Child Protection cases.

early intervention and therefore avoid a referral to Social Care. With regards to limitation and urgent protection, there were examples of cases where GCP2 provided evidence to escalate a case or where the evidence was used in legal proceedings.

Participants generally felt the grading constructs were well defined, with some describing them as “thorough and accurate” and “easy to use”. However, some felt that they could be too narrow or prescriptive in parts (particularly grade 5), with not enough consistency between grades or would elicit very similar information (for example, reciprocation and quality of the relationships between parent and child, response timing and sensitivity). These individuals felt that the grading constructs would benefit from further improvement (see Appendix D2).

Responses were positive with regards to the extent to which GCP2 enables intervention to be targeted. Participants explained that the tool includes clear recommendations with a timescale, pinpoints what support is required and enabled resources to be targeted – although this is subject to resources being available locally. One response however, was less positive and referred to the lack of consideration of parental problems:

“...there is not enough about the nature of the parents’ problems to enable an analytic assessment to take place which would identify the nature of the difficulties [often many in neglect cases] and appropriate targeting of intervention.”

Neglect expert

There was some doubt with regards to the extent to which GCP2 enables a child's development or developmental disorder to be assessed, and the extent to which it enables a professional to understand educational, emotional or behavioural outcomes in a child<sup>10</sup>. There was a sense that these were not necessarily an integral part of GCP2, for example assessment of development or developmental disorder was seen as a "specialist health assessment". Respondents felt that the grading constructs generally do not facilitate this, with such information usually being obtained from other professionals during the course of an assessment. Any information picked up in relation to these issues during the GCP2 assessment could also then inform other assessments, for example GCP2 could help to identify a "potential" development disorder, in which case the social worker would seek expert opinion/assessment.

Encouragingly, all participants felt that GCP2 measures what it intends to measure to a large extent. One summarised:

"Yes I think the GCP2 provides clear, evidence based information to assist local authorities to make decisions about a child's care, needs and any plans necessary."

Practitioner

Participants were also very positive regarding the extent to which GCP2 is consistent with current knowledge and best practices in child neglect.

---

10 These are listed under the purpose of the GCP2, page 7 of the GCP2 guidance (Srivastava and Hodson, 2014).

## Discussion

### Summary

The evidence suggests that the GCP2 is a valid and reliable measure of child neglect which has:

- moderate to high inter-rater reliability (with strong internal consistency)
- a high level of concurrent validity
- a fairly high level of face validity

### Reliability

Agreement was statistically significant for all teams collectively across all domains, but not for each team individually for each domain. The lowest level of agreement was from Team A ratings of Care of Safety. There are two possible explanations for this:

- This team typically received referrals for large families. This meant that it was not always possible for the two practitioners to stay together during visits, resulting in one practitioner witnessing safety issues which were not seen by the co-rater. This emerged in the moderation exercise, after GCP2 scores were submitted.
- Team A was the first team to start using the GCP2 for the testing process. In one family in particular, one rater included the garden in the scoring of section B while the other did not. This resulted in some discrepancy in the scoring, as the garden was very unsafe and “worse than the state of the house”. However, once this issue had become apparent, all participating practitioners were advised that the garden should be included in scoring, thereby improving the inter-rater reliability for Area B during the course of the testing.

## Validity

**Concurrent validity:** GCP2 correlates very well with existing tools which measure aspects of neglect. There were only a small number of items where the correlation was not statistically significant: four of the sub-areas from Area A did not correlate strongly with some of the related areas from other tools. These were mostly items related to health, and are likely to be due to the fact that the descriptions/ definitions of these items are not an exact conceptual match between tools. For example, part of the criteria for judging the Physical Health item in NCFAS-G relates to the child's physical health status, whereas this is not included in the GCP2. In addition, 'environmental risks' in NCFAS-G refers to the "family's living and neighbourhood circumstances" whereas this was compared to GCP2's housing domain which only refers to the home and not the neighbourhood.

**Face validity:** The GCP2 can be said to have a fairly high level of face validity, with participants generally agreeing that the GCP2 achieves what it sets out to do. However, the results suggest that it may be beneficial to review the extent to which the GCP2 aims to enable the assessment of child development and developmental disorder, as well as educational, emotional and behavioural outcomes, particularly how the impact of parental problems on children are reflected in the scoring.

## Limitations

Limitations centre on the representativeness of the sample. All of the cases were already identified as having 'children in need' or 'child protection' status, hence the sample does not include cases at a much earlier stage of need identification. However,

as stated in the results section there was a range of scores, so the sample includes cases with a range of types and levels of concern. The vast majority of children were White British, so testing did not include a range of ethnicities, thus limiting generalisability. In addition, the validity testing only included three children under five. These demographic limitations could be addressed in any future psychometric testing.

It is possible that the NSPCC practitioners carried out more family visits than may typically take place when GCP2 is used by a Local Authority. There is therefore a potential limitation in generalising to different contexts. However, the decision was taken to carry out testing within NSPCC for a number of reasons, in particular because this provided the opportunity to carry out a more robust test of the measure. In order to achieve optimal inter-rater reliability testing, both raters in a pair should have the same level of knowledge about the family so that prior knowledge does not skew the scoring. It would have been more difficult to achieve this within a Local Authority, where a lot of cases would typically have an assigned key worker who could be familiar with the family. In comparison, when cases were referred from the Local Authority to the NSPCC, both NSPCC practitioners had no prior knowledge of the case, with one exception, and therefore all scoring was determined by evidence collected together, resulting in a stronger test of the tool.

## Testing in a practice context

This testing was carried out in a service setting, in which families were receiving assessment and intervention, as appropriate. As such, the process involved navigating a tension between practice imperatives and the requirements of reliability and validity testing. These tensions were addressed and agreements reached in ways that were comfortable for the practitioners in their practice but which did not undermine the robustness of the testing process. There were two issues that were particularly notable.

The first related to the fact that practitioners would usually carry out an assessment in relation to all children in a family where there are concerns about neglect. However, as explained above, the testing process required cases which could be considered as reasonably distinct cases. There was therefore a process whereby the requirements of the testing were considered in relation to the practice requirements to ensure all children's needs were assessed, as appropriate. This involved discussions with practitioners, managers, and the team carrying out the testing. It was agreed that GCP2 would be carried out on all children who warranted assessment, and that all GCP2 scores would be submitted, with a maximum of two cases per family being used for the analysis. This also meant that each team needed to recruit five families (there were at least two cases in each family), which was achievable in realistic timescales. Therefore the position reached was satisfactory to all parties.

There was also a negotiation process in relation to the level of discussion that could take place between two practitioners in a pair, and the timing of this. In order to meet the requirements for testing, each practitioner's scoring should not be influenced by

their co-rater. In order to achieve this, reflection and discussion in relation to the evidence gathered should not take place prior to scoring. Practitioners felt, however that this contravened good practice guidelines that they followed, of shared reflection and discussion after a visit. As with the other issue, there was discussion with all parties, and the requirements of both good social work practice and robust reliability testing were considered by all. It was agreed that the scoring would be carried out at the earliest opportunity, usually after each visit or when there was enough information to score each domain. Practitioners were then able to reflect on the area(s) that had been scored. This approach worked in practice.

### Comparability to original testing

The level of agreement in the current inter-rater reliability testing was slightly lower than in the original testing (Srivastava and Polnay, 1997). However it is unlikely that this indicates that GCP2 has lower inter-rater reliability than the original GCP. It is difficult to make direct comparisons between the two sets of testing for a number of reasons, including that the tools were implemented differently. In the original testing, scoring was based on one short visit (average 25 minutes). However in the current testing the evidence was collected over a number of visits, and from different sources including conversations with parents, feedback from children where appropriate, observations in the home and other professionals. In addition, some unannounced visits took place as required. Therefore, with a greater volume and range of evidence, it would seem likely that there was more scope for differing views and different interpretations of the evidence. Considering this, it is encouraging

that the level of agreement is generally relatively high in the current testing.

In addition the demographics of the samples differed greatly, notably in terms of the level of concerns and child age groups. It is possible that this sort of tool is implemented differently according to such demographics, and therefore that this could influence the level of inter-rater reliability.

## Conclusion

GCP2 has been found to be reliable and valid. It can be used in the knowledge that it has sound psychometric properties, and is a reliable and valid assessment tool in aiding practitioners in the assessment of child neglect.

# References

- Barlow, J., Fisher, J. and Jones D. (2012) Systematic review of models of analysing significant harm DFE-RR199
- Caldwell, B. M., and Bradley, R. H. (2003). Home Observation for Measurement of the Environment: Administration Manual. Tempe, AZ: Family & Human Dynamics Research Institute, Arizona State University.
- George, D., and Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- Johnson, R and Cotmore, R. (2015) National evaluation of the Graded Care Profile, NSPCC
- Kirk, R. S., and Ashcraft, K. R. (1998). User's guide for the North Carolina Family Assessment Scale (NCFAS). Chapel Hill, NC: University of North Carolina.
- Kirk, R. and Martens, P. (2006) Development and Field Testing of the North Carolina Family Assessment Scale for General Services (NCFAS-G), NFPN.
- Landis JR and Koch GG. The measure of observer agreement for categorical data. *Biometrics* 1977;**33**:159–74.
- Phelan, C. and Wren, J (2006) Exploring reliability in academic assessment: <https://www.uni.edu/chfasoa/reliabilityandvalidity.htm> Accessed 15/9/15
- Srivastava, O.P. and Hodson, D. (2014) Graded Care Profile version 2 Guidance

Srivastava, O.P., Polnay, L. (1997) Field trial of graded care profile (GCP) scale: a new measure of care. *Archives of Disease in Childhood* 1997;76:337–340

Totsika, V. and Sylva, K. (2004). The Home Observation for Measurement of the Environment Revisited. *Child & Adolescent Mental Health Journal*, 9 (1), 25–35.

# APPENDICES

## Appendix A: Evidence for use of other tools

### Home Inventory

Research has demonstrated HOME's sound psychometric properties. A review by Oxford University concluded that "HOME is without doubt the most commonly used environmental assessment instrument in developmental research. Many years of research have demonstrated the important correlations it has with measures of cognitive and language development and its ability to independently predict such outcomes later in the child's life. Most importantly, however, research has proved the instrument's validity in describing the home environments of children at risk and revealing the effect of home experiences in developmental outcomes." (Totsika & Sylva, 2004)

### The North Carolina Family Assessment Scale for General Services (NCGAS-G)

This scale was developed from the original NCFAS (North Carolina Family Assessment Scale) so that it could be used with lower risk as well as high risk families in a range of practice settings. NCFAS-G was found to be a very reliable measure with good psychometric and statistical properties. Internal consistency was described as very good "with respect to the DR workers assessing low-moderate risk families in the DR practice environment"). The results of the assessments using this measure with a sample of service users were as expected based on the population for which it was designed (Kirk, 2006).

# Appendix B: GCP2/Ncfas-G/ Home Inventory mapping

GCP2	NCFAS-G	HOME inventory*
<b>Area of Physical care</b>	D6. Absence/ presence of neglect of children	
A1. Nutrition	G4. Food and nutrition H4. Children's physical health C6. Family routines/ rituals	
A2. Housing	A4. Environmental risks A5. Habitability of housing	
A3. Clothing	A6. Personal Hygiene	
A4. Hygiene	A6. Personal Hygiene	
A5. Health	H4. Children's physical health H7. Family access to health/ mental health care	
<b>Area of Care of Safety</b>	D6. Absence/ presence of neglect of children	A total B total
B1. Safety In Parents Presence	D7. Absence/ presence of access to weapons A3. Environmental risks A4. Habitability of housing B6. Controls access to media/ reading material	
B2. Safety when parent is absent	B1. Supervision of children	
<b>Area of Emotional Care</b> (Nature of attachment)	D6. Absence/ presence of neglect of children	A total B total
C1. Carer's responsiveness	C1. Bonding with children C2. Communication with children	A total B total
C2. Mutual engagement	C1. Bonding with children C2. Communication with children E3. Children's relationship with parents/ caregivers	A total B total
<b>Area of Developmental Care</b>	D6. Absence/ presence of neglect of children	A total B total
D1. Stimulation	A7. Learning environment B1. Supervision of children B3. Provision of developmental/ enrichment opportunities B5. Promotes children's education B7. Parents/caregivers literacy C6. Family routines/ rituals C7. Family recreation and play activities E2. School performance F2. Relationships with child care, schools and extracurricular activities E5. Children's relationship with peers	A total B total
D2. Approval	B2. Disciplinary Practices D4. Absence/ presence of emotional abuse of Children	A total B total

D3. Disapproval	B2. Disciplinary Practices	A total
	C1. Bonding with children	B total
	C2. Communication with children	
	D3. Absence/ presence of Physical abuse of Children	
	D4. Absence/ presence of emotional abuse of Children	
D4. Acceptance	C1. Bonding with children	A total
	C3. Expectations of children	B total

\*Please note 'A total' refers to the domain total for Responsivity and 'B total' refers to the domain total for Acceptance (or Emotional Climate), depending on the child age group.

# Appendix C: Additional Tables

Table 6: Level of concern by site

	Site A	Site B	Site C	Total
<b>Potential concerns</b>	0	0	0	0
<b>Child in Need</b>	2	2	8	12
<b>Child Protection</b>	8	8	2	18
<b>Total</b>	10	10	10	30

Table 7: Kappa scores for sub areas (for all teams combined)

	Kappa	P value
A. Area of Physical care		
A1. Nutrition	.318	< 0.05
A2. Housing	.398	< 0.01
A3. Clothing	.699	< 0.01
A4. Hygiene	.522	< 0.01
A5. Health	.885	< 0.01
B. Area of Care of Safety		
B1. Safety In Parents Presence	.325	< 0.01
B2. Safety when parent is absent	.839	< 0.01
Area of Emotional Care		
C1. Carer's responsiveness	.435	< 0.01
C2. Mutual engagement	.484	< 0.01
Area of Developmental Care		
D1. Stimulation	.582	< 0.01
D2. Approval	.494	< 0.01
D3. Disapproval	.391	< 0.01
D4. Acceptance	.525	< 0.01

Table 8: Correlations between GCP2 Area D subscales and related NCFAS-G subscales and HOME total scales

NCFAS-G/HOME sub scales/total scales	GCP2 sub areas			
	Stimulation	Approval	Disapproval	Acceptance
Learning environment (NCFAS-G)	- 0.83**	-	-	-
Supervision of children (NCFAS-G)	- 0.60*	-	-	-
Provision of development opportunities (NCFAS-G)	- 0.77**	-	-	-
Promotes children's education (NCFAS-G)	- 0.81**	-	-	-
Parent/caregiver's literacy (NCFAS-G)	- 0.68**	-	-	-
Family routines/rituals (NCFAS-G)	- 0.67**	-	-	-
Family recreation and play activities (NCFAS-G)	- 0.80**	-	-	-
School performance (NCFAS-G)	- 0.74**	-	-	-
Relationships with childcare (NCFAS-G)	- 0.61*	-	-	-
Children's relationships with peers (NCFAS-G)	- 0.85**	-	-	-
Disciplinary practices (NCFAS-G)	-	- 0.75**	- 0.78**	-
Bonding with children (NCFAS-G)	-	-	- 0.81**	- 0.85**
Communications with children (NCFAS-G)	-	-	- 0.78**	-
Expectations of children (NCFAS-G)	-	-	-	- 0.82**
Absence/presence of emotional abuse (NCFAS-G)	-	- 0.80**	-	-
Absence/presence of physical abuse (NCFAS-G)	-	-	-0.84**	-
A Total (HOME)	- 0.74**	- 0.74**	- 0.82**	- 0.81**
B Total (HOME)	- 0.58*	- 0.65*	- 0.69**	- 0.78**

\*Spearman correlation coefficient is significant at the 0.05 level, \*\*Spearman correlation coefficient is significant at the 0.01 level. GCP2 = Graded Care Profile 2nd version, NCFAS = North Carolina Family Assessment Scale for General service, HOME = Home Inventory Record Sheets.

+

Table 9: Correlations between GCP2 Area A subscales and related NCFAS-G subscales

NCFAS-G subscales	GCP2 subscales				
	Nutrition	Housing	Clothing	Hygiene	Health
Food and nutrition	- 0.77**	-	-	-	-
Children's physical health	- 0.21	-	-	-	- 0.16
Family routines/rituals	- 0.67**	-	-	-	-
Environmental risks	-	- 0.31	-	-	-
Habitability of housing	-	- 0.54*	-	-	-
Personal Hygiene	-	-	-0.73**	-0.86**	-
Access to health care	-	-	-	-	- 0.47

\* Spearman correlation coefficient is significant at the 0.05 level, \*\*Spearman correlation coefficient is significant at the 0.01 level. GCP2 = Graded Care Profile 2nd version, NCFAS-G = North Carolina Family Assessment Scale for General Services.

Table 10: Correlations between GCP2 Area B subscales and related NCFAS-G subscales

NCFAS subscales	GCP2 subscales			
	Environmental risks	Habitability of housing	Controls/ access to media	Supervision of children
Safety in parents' presence	- 0.85**	- 0.71**	- 0.68**	-
Safety when parent is absent	-	-	-	- 0.82**

\* Spearman correlation coefficient is significant at the 0.05 level, \*\*Spearman correlation coefficient is significant at the 0.01 level. GCP2 = Graded Care Profile 2nd version, NCFAS-G = North Carolina Family Assessment Scale for General Services

Table 11: Correlations between GCP2 Area C subscales and related NCFAS-G subscales and HOME total scales

NCFAS-G/HOME subscales/ total scales	GCP2 subscales	
	Carer's responsiveness	Mutual engagement
Bonding with children (NCFAS-G)	- 0.80**	- 0.85**
Communication with children (NCFAS-G)	- 0.86**	-
Children's relationship with parents/carers (NCFAS-G)	-	- 0.81**
A Total (HOME)	- 0.79**	- 0.82**
B Total (HOME)	- 0.74**	- 0.76**

\*Spearman correlation coefficient is significant at the 0.05 level, \*\*Spearman correlation coefficient is significant at the 0.01 level. GCP2 = Graded Care Profile 2nd version, NCFAS-G = North Carolina Family Assessment Scale for General Services. HOME = Home Inventory Record Sheets.

# Appendix D: Face validity<sup>11</sup>

## 1) Further suggestions for missing areas

- *Educational neglect, the use of the word 'schooling' should be replaced mostly with education in D1 Stimulation 5+ yrs section and the text should add in issues like helping with homework, valuing education, encouraging achievement and actual school performance. Non-attendance at school should also be added as a must.*
- *Achievement of language skills (related to stimulation)*
- *Under A5 Health, it is too focussed on physical illness. It should also mention mental health appointments for children [including help with their emotional and behavioural problems].*
- *Under Housing there should be a coding/reference to the house being cold or poorly heated and also to adequate bedding.*
- *Other issues you might want to consider include anti-social behaviour being modelled to a child; child not protected from harm by the other parent or another adult; expectations age-appropriate or not age-appropriate; whether parents work together /are consistent in their approach to the child [e.g. if mother good but father victimises the child - needs to be caught]; parents' acceptance of help; parental ambivalence.*
- *There needs to be a proper space - perhaps at the top and in the template- about child's problems as above and including e.g. autism, ADHD, health etc.*

---

11 All data in Appendix D comes directly from the neglect expert

## 2) Grading constructs

- *Sometimes the wording of the codes, especially for 5, is too prescriptive and so may not be used correctly. e.g. Online safety under 5 says 'despite understanding the danger' which hasn't appeared in 1-4 so a parent might have careless disregard and not understand the danger so people won't code as 5. Suggest 'despite understanding the danger' is removed, ....some of the descriptions... are not quite in keeping with the other descriptions and become too narrow.*
- *Also, e.g. Disapproval measure the grades are about how often, and don't include a notion of severity although it is implied. Also much of the time might be better than all of the time for 5.*



