

ANNA GOOD AND JOY MACKEITH *Triangle Consulting, The Dock Hub, Wilbury Villas, Hove, East Sussex, England*

Assessing Family Functioning: Psychometric Evaluation of the Family Star Plus

Objective: *To examine the psychometric properties of the Family Star Plus, an assessment of family functioning using a collaborative mode of completion between the caseworker and service user.*

Background: *The Family Star Plus is widely used to assess strengths and needs within services supporting families.*

Method: *In Phase 1, interrater reliability was assessed. Phase 2 established factor structure, internal consistency, item redundancy, and responsiveness. In Phase 3, the Family Star Plus was used to predict children's unauthorized school absence.*

Results: *Workers reliably applied Family Star Plus scales. A unidimensional factor structure was found with good internal consistency and responsiveness. School absence was predicted by the Family Star Plus.*

Conclusions: *These findings demonstrate the validity of the Family Star Plus as a tool for assessing needs and measuring progress of family support interventions.*

Implications: *The Family Star Plus can be confidently used and recommended as an outcomes measurement tool.*

In efforts to be more accountable and show demonstrable change following intervention programs, service providers have adopted more

routine use of outcomes measures to demonstrate intervention successes, with at least three quarters of not-for-profit organizations measuring all or some of their work (Ógáin et al., 2012). Policymakers value outcome measurement of family support, health, social, and education services (Hoffman et al., 2006) and are increasingly focused on identifying services and interventions that can be shown to improve children's lives (Klett-Davies et al., 2009). Interest in outcome assessments is a particular focus in countries where there is clear indication of a need to further support families. For instance, the interest in outcomes assessment is fuelled by the poor state of child well-being in the United Kingdom and United States relative to other affluent nations (Bradshaw et al., 2007a, 2007b; Collishaw et al., 2004) and the subsequent societal costs of poor child outcomes (Scott et al., 2001).

In the United States, measurement of family outcomes is mandatory for early intervention programs (Early Childhood Outcomes Center, 2005). In the United Kingdom, the government's Troubled Families Initiative offers payment-by-results support based on outcomes related to family functioning (e.g., crime, education, antisocial behavior, and progress to work) to improve outcomes for families with entrenched and complex problems (Department of Communities and Local Government [DCLG], 2012). To meet assessment demands of a payment-by-results framework, more than 50 local councils in the United Kingdom use the Family Star Plus assessment (Burns & MacKeith, 2017; DCLG, 2016) as part of their

Triangle Consulting, The Dock Hub, Wilbury Villas, Hove, BN3 6AH, United Kingdom
(anna@triangleconsulting.co.uk).

Key Words: Evaluations, family support, parent-child relationships, parent education and guidance, quantitative.

delivery and of payment-by-results submissions within the Troubled Families Initiative.

FAMILY STAR PLUS

The Outcomes Star is a suite of tools created by Triangle Consulting Social Enterprise both to support and measure change when working with people. The Outcomes Star tools not only measure outcomes but are also instrumental in helping the desired outcomes to be achieved (Arvidson & Kara, 2013). The tools facilitate conversations about strengths, needs, and progress between caseworkers and service users.

There are more than 30 versions tailored to particular sectors, contexts, and client groups, with each version created through an iterative process of data gathering (workshops with managers, frontline workers, and service users and literature reviews), drafting of scales and further refinement (see MacKeith, 2011). The Family Star Plus is one version of the Outcome Star (Burns & MacKeith, 2017), developed from the Family Star, an earlier version (Burns & MacKeith, 2013). The Family Star Plus assessment was created to meet a need for a casework and outcomes measurement tool for services aimed toward improving family functioning and children's well-being and life chances.

All versions of the Outcomes Star consist of a number of "outcome areas" (e.g., Physical Health, Your Well-Being) arranged in the shape of a star, with each outcome area within a Star underpinned by detailed descriptions for each point on a five-step "Journey of Change" scale. For the Family Star Plus, the five Journey of Change steps are broken down into Stuck (1–2), Accepting help (3–4), Trying (5–6), Finding what works (7–8), and Effective parenting (9–10). Service users and caseworkers discuss all the areas of the service user's life represented on the Star and agree where they are on the Journey of Change (Burns & MacKeith, 2013).

The collaborative process of completing the Star facilitates its use as a case-work tool, making use of the knowledge and judgment of both service users and caseworkers and giving service users "a voice within the assessment process" (DCLG, 2016, p. 56). In some situations, caseworkers may challenge service users' perceptions based on their knowledge of the assessment steps or professional judgment, but separate "worker only" and "service use only" readings are recorded when agreement cannot be

reached. Once readings are agreed on, they are plotted on the Star diagram. When the process is repeated (usually every 3–6 months), the difference in the two readings provides a picture of change.

Existing Research on the Suite of Outcomes Stars

There is a growing body of literature evidencing the value of the Outcomes Star suite of tools for caseworkers and for measuring outcomes. The suite of Outcomes Star tools has been described as providing an improved quality of case management on the basis of feedback from caseworkers and service users (Harris & Andrews, 2013). Focusing specifically on the Family Star, an independent evaluation including analysis of more than 3,000 completed Stars and interviews with service users, frontline staff and managers concluded that it contributes to developing parental resilience, facilitates reflective casework practice, and provides valuable insights into positive change and areas needing further investigation (York Consulting, 2013).

Triangle, the creators of the Outcomes Star suite of tools, conducts ongoing psychometric evaluation work, including at the development stage. There are independent validation articles for a number of versions of the Outcomes Star (e.g., Bailey & Kerlin, 2015; Dickens et al., 2012; Griffiths et al., 2015; Killaspy et al., 2012; Larsen & Griffiths, 2013; Placentino et al., 2017). For example, Placentino et al. (2017) reported that the Recovery Star, a version of the Outcomes Star for individuals experiencing mental health difficulties, has good test-retest reliability, convergence with other validated tools, and inter-rater reliability.

Current Research

This research looks at the psychometric properties of the Family Star Plus. The Family Star Plus focuses on aspects of family functioning that relate to effective parenting. As far as we are aware, the current research is the first to report on the psychometric properties of the Family Star Plus. However, initial support for interrater reliability when using the Family Star is reported by MacKeith (2014). This assessment reports outcomes from a small sample of caseworkers who assigned Journey of Change readings (using the 1–10 scale) to a written service user case study.

Table 1. *The 10 Areas of the Family Star Plus*

-
1. **Physical Health.** Taking children for treatment when they have health problems and giving them a healthy lifestyle
 2. **Your Well-Being.** Parent's well-being, mental health, and ability to deal with problems
 3. **Meeting Emotional Needs.** Being able to meet children's emotional needs and having a positive parent-child relationship
 4. **Keeping Your Children Safe.** Protecting children including avoiding accidents, bullying, and abuse
 5. **Social Networks.** Parents and children having the social contact they need
 6. **Education and Learning.** Supporting children's learning and aspirations
 7. **Boundaries and Behaviour.** Parents providing appropriate boundaries and consequences and being a positive role model
 8. **Family Routine.** Having an appropriate weekday routine including getting to school on time and having regular mealtimes
 9. **Home and Money.** Having a stable and adequate home; managing money so that the family's needs are met
 10. **Progress to Work.** Moving toward parents being employed
-

The current study is structured into three phases. In Phase 1, interrater reliability was assessed. Phase 2 used routinely collected data to establish factor structure, internal consistency, item redundancy, and responsiveness to change. In Phase 3, a second data set was used to test the predictive validity of the Family Star Plus. This phase involved predicting unauthorized school absence from readings on two relevant Star areas (Education and Learning; Boundaries and Behaviour). The Education and Learning area was expected to have the most direct association with school absence as the scale directly refers to school attendance and supporting learning and aspiration more generally. Time 1 scores on the Boundaries and Behaviour area were also expected to show an association with school attendance because it measures whether children have appropriate boundaries and rules in place.

PHASE I. INTERRATER RELIABILITY

Method

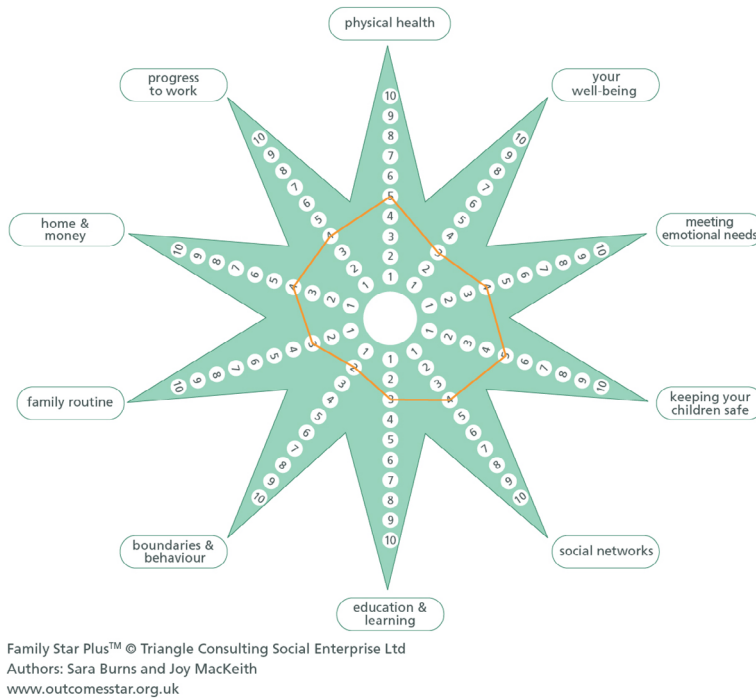
Participants. The participants were an opportunity sample of caseworkers in early help teams (teams providing support to families when problems first emerge), within two county councils in the United Kingdom ($N = 65$; 20% of all caseworkers in the two councils). All caseworkers in specific teams were invited to participate as part of their normal working practice. Most of the participating caseworkers were female (85%). Caseworkers had at least level 3 National Vocational Qualifications, a work competence-based qualification indicating competence in application of knowledge and skills in a broad range of

varied work activities performed in a wide variety of contexts and most of which are complex and nonroutine. A level 3 qualification status entails "considerable responsibility and autonomy, and guidance of others is often required" (Nijhof & Streumer, 2012, p. 116).

A short time before being recruited, all caseworkers had completed a 1-day training course designed to familiarize them with the detailed scales of the Family Star Plus and how to use the Star in casework. At the end of the training, all caseworkers' understanding was informally checked through group discussion around service user stories (i.e., short paragraphs describing the service user's situation on each of the 10 outcomes on the Family Star Plus). The 10 outcome areas of the Family Star Plus are shown in Figure 1 (see Table 1 for details). Reviews of worker understanding and use of the scales was conducted also in subsequent supervision meetings.

Family Star Plus Stories for Scoring. A real story of a service user in an early help team was used to create a written profile describing the service user's situation for each of the 10 outcome areas of the Family Star Plus. The resulting Readings (i.e., short paragraphs and related scores on the Journey of Change for each of the outcome areas) were agreed on by five trainers at Triangle Consulting, the creators of the Outcomes Star, who were experienced in training people to use the Outcomes Star. To establish these correct readings, each of the five members of Triangle staff independently assigned a score between 1 and 10 on

Figure 1. THE FAMILY STAR PLUS STAR DIAGRAM.



Note. This image is reproduced with permission from Triangle, the creators of the Outcomes Star.

the Journey of Change using the short description for each outcome area. Revisions were made until agreement on rating of readings for each outcome area were agreed on by all trainers. These agreed-on readings were used to evaluate caseworker accuracy in applying the Journey of Change descriptors.

Study Procedure. To test interrater reliability of caseworkers' scores for readings, caseworkers were asked to independently give a reading (a score between 1 and 10) for each of the 10 outcome areas in the profile using the descriptors for each Journey of Change stage in the Family Star Plus User Guide.

In this task designed to assess interrater reliability, caseworkers assigned readings based on the sort of information they would normally extract through conversations with service users but without the normative practice of discussing the stories and potential scores with other caseworkers. For this study, caseworkers gave a reading between 1 and 10 for each of the 10 outcome areas in the Family Star Plus user guide. The 10-point continuum for each of the

10 outcomes areas reflects a five-stage journey: Stuck (1–2), Accepting help (3–4), Believing and trying (5–6), Finding what works (7–8), and Effective parenting (9–10). Each point on the scales is clearly defined, for example, for Physical Health area the Stuck stage includes: “People say I am neglecting my children’s health. I don’t want to talk about it: You don’t take your children to the doctor or dentist except in emergencies. Your children don’t eat healthy food, have too little exercise or sleep, and may be obese, undernourished, or at risk of health problems later ... Choose 1 if you don’t see this as a problem. Choose 2 if you have moments of concern over your children’s health but don’t accept support so there is no change.” Caseworkers were supervised by their managers to ensure that each caseworker assigned readings the scores independent of input from other caseworkers. Reliability was determined by examining caseworkers’ responses to establish how accurately each worker had assigned readings to the areas of case study in relation to the predetermined answer key determined by the trainers, as well as agreement with the other caseworkers.

Analytic strategy. An SPSS macro was used to calculate Krippendorff's alpha for ordinal data (Hayes & Krippendorff, 2007). Krippendorff's alpha assesses disagreements as well as agreements, and can be used with any number of observers, nominal, ordinal, interval, and ratio data, with or without missing data (Hayes & Krippendorff, 2007). Values of Krippendorff's α range from 0.00 for absence of reliability to 1.00 for perfect reliability, with .80 and higher considered necessary for high stakes decision making and values between .67 and .80 allowing tentative conclusions to be drawn (Krippendorff, 2004). In the current study, Krippendorff's alpha was calculated for each caseworker (in relation to the predetermined answers), and a mean value was obtained across the whole cohort of 65 caseworkers. In addition, we calculated Krippendorff's alpha for consistency between caseworkers.

RESULTS

The values of Krippendorff's α for the caseworkers ranged from .39 to 1.0, with a mean of .83 indicating very good interrater reliability and the same mean for male and female caseworkers (.83). The alpha coefficient was above .80 for 44 caseworkers (68%) and less than .67 for only nine of the 65 caseworkers (14%). The small sample size meant statistical comparisons of these nine caseworkers to the rest of the cohort was not possible, but they had a similar ratio of males to females compared to those with a coefficient above .67 (89% vs. 84% were female) and had completed slightly more Star readings ($M = 7.88$, $SD = 3.95$ vs. 5.55 , $SD = 4.42$). The alpha coefficient for consistency between caseworkers was .75.

PHASE 2. ANALYSIS OF ROUTINELY COLLECTED FAMILY STAR PLUS DATA

Method

Participants and procedure. Workers at the county councils participating in Phase 1 entered routinely collected, deidentified Family Star Plus data into the Star Online database for all parents who completed Stars between February 2013 and December 2015. The data comprised "Star readings," that is, scores on the 1 to 10 Journey of Change for each of the 10 outcome areas for each service user. For inclusion in the

analyses for this study, each service user had to have at least two Star readings, resulting in information for 948 service users. Second readings were completed at the end of service use for some families, but often service users continued to receive support and complete further Star readings. Scores on the 10 outcome areas at Time 2 are compared with those from Time 1.

Most service users were White British ($n = 653$, 75%), female ($n = 572$, 62%), and of working age ($n = 918$, 98%). All were using a "Tier 2" early help service. These services offer the first level of early or specialist help to families in which children have additional needs and may require extra support to promote their welfare and prevent their needs from becoming more complex. Service users were offered targeted services, such as school counseling, parenting programs, and support for teenage parents.

There was not a set point at which Family Star Plus readings were completed because this is tailored to the needs of individual service users. Forty-five percent of first Stars were completed on the day of service entry ($M = 44.82$ days, $Mdn = 1$ day, $SD = 125.95$, range = 0–153, skew = 5.52, kurtosis = 40.27). Of the second Stars, 80% were completed within 6 months of first Stars ($M = 126.76$ days, $Mdn = 108.5$, $SD = 82.14$, range 7–500, skew = 1.34, kurtosis = 2.36). Almost all Star readings were completed in collaboration with a caseworker (94% of first Stars and 87% of second Stars). There was no missing Family Star Plus data because the online system requires readings to be entered for each outcome area.

Analytic strategy. The expected unidimensional structure of the Family Star Plus assessing family functioning was tested using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), with the sample of Time 1 Family Star Plus readings (i.e., scores on each of the 10 outcome areas) from 948 service users were randomly assigned to the calibration ($n = 474$) and validation sample ($n = 474$). The EFA was performed using parallel analysis based on minimum rank factor analysis (Timmerman & Lorenzo-Seva, 2011) with an oblique rotation (Promin) within FACTOR version 10.8.02 (Lorenzo-Seva & Ferrando, 2006). Reviews of previous studies suggest that parallel analysis is one of the most accurate factor-retention methods (e.g., Fabrigar et al., 1999; Hayton

et al., 2004; Henson & Roberts, 2006). It is based on polychoric correlations, appropriate for the ordinal level of measurement used here (Garrido et al., 2013; Ruscio & Roche, 2012). The CFA was performed using the validation sample within the AMOS program (Arbuckle, 2014). Because there was multivariate nonnormality, the unweighted least squares method was appropriate (Kline, 2011). The root mean square error of approximation can only be calculated in AMOS when multivariate normality is assumed, so model fit was assessed using the comparative fit index ≥ 0.90 and the standardized root mean square residual ≤ 0.08 (Hu & Bentler, 1998; Schermelleh-Engel et al., 2003). IBM SPSS 24.0 was used for the remaining analyses, which included the use of Cronbach's (1951) alpha as an indicator of internal consistency, interitem correlations to assess redundancy through repetition, and the Wilcoxon signed rank test for paired scores to assess the extent to which service users showed change in each of the 10 outcome areas. This test is the equivalent of a repeated-measures *t*-test for ordinal data.

RESULTS

Factor Analysis, Item Redundancy, and Internal Consistency

The suitability of the data for the exploratory factor analysis was confirmed with a Kaiser–Meyer–Olkin (KMO) value of 0.89, exceeding the recommended minimum value of .60 (Kaiser, 1970) and a significant Bartlett's test of sphericity, $\chi^2(45) = 157.90$, $p < .001$ (Bartlett & Jackson, 1954). Polychoric correlations were used based on the ordinal level of the data and kurtosis within the data. In parallel analysis, eigenvalues in randomly simulated data are compared with those of the factors in the actual data set and the number of factors at the point where the eigenvalue in the simulated data is greater than that of the actual data is considered significant. On this basis, a unidimensional factor structure was advised, explaining 68.70% of the variance, with an eigenvalue for the first factor of 4.59. The largest factor loadings were for Physical Health, Meeting Emotional Needs, and Family Routine (all $>.70$) and weakest for parental Progress to Work (.51). Internal consistency was very good (Cronbach's $\alpha = 0.86$), and no interitem

correlations exceeded .70, indicating that there was no item redundancy.

Before conducting CFA on the validation sample, univariate and multivariate normality were checked. None of the individual indicators departed significantly from a univariate normal distribution, but Mardia's coefficient was 10.99 indicating multivariate non-normality (Yuan et al., 2002), so the unweighted least squares method was selected. The Goodness of Fit Index was .986, indicating that approximately 99% of the covariance matrix could be explained by the single factor model (Kline, 2011). The SRMR was .058 (smaller than the cutoff threshold of $\leq .08$), confirming the model suggested by the EFA.

Responsiveness. As can be seen in Table 2, the signed rank test revealed a statistically significant increase in scores for all areas of the Family Star Plus ($p < .001$) between Time 1 and Time 2, with medium effect sizes for Your Well-Being, Meeting Emotional Needs, Social Networks, and Education and Learning and small–medium effect sizes for the remaining five areas. At least one third of service users improved in each area using the full sample, which included those who began at 10 and therefore could not move forward. For all outcome areas, the range was 1 to 10 and skewness and kurtosis values were well within the accepted range of -2 and $+2$ (George & Mallery, 2010). When we restricted the sample to those with the greatest needs (at 1–6 on the individual outcome areas), the effect size was large for four areas and medium–large for the remaining six ($p < .001$, See Table 3).

PHASE 3. TESTING THE PREDICTIVE VALIDITY OF THE FAMILY STAR PLUS

Method

Participants and procedure. Time 1 readings (i.e., scores on the 1–10 Journey of Change) on relevant areas of the Family Star Plus (Education and Learning and Boundaries and Behaviour) were used to predict the percentage of the total required school days missed using school systems for 388 children of parents who completed the Family Star Plus within a United Kingdom county council. The Education and Learning area is about parents supporting their children's learning and aspirations, with stages ranging from Stuck ("I am not thinking about

Table 2. Responsiveness of the Family Star Plus for All Service Users (N = 948)

| Scale | Time 1Mdn (IQR) | Time 2Mdn (IQR) | Wilcoxon Z | Effect size ^a r | Readings improved (%) | Skew | Kurtosis |
|----------------------------|-----------------|-----------------|------------|----------------------------|-----------------------|------|----------|
| Physical Health | 7 (6–9) | 8 (7–9) | 12.24*** | 0.28 | 48 | –.52 | –.44 |
| Your Well-Being | 6 (4–7) | 7 (6–9) | 16.48*** | 0.38 | 62 | –.08 | –.65 |
| Meeting Emotional Needs | 7 (5–8) | 8 (6–9) | 15.26*** | 0.35 | 33 | –.21 | –.36 |
| Keeping Your Children Safe | 8 (6–9) | 9 (7–10) | 11.34*** | 0.26 | 46 | –.62 | –.35 |
| Social Networks | 7 (5–8) | 8 (6–9) | 15.55*** | 0.33 | 55 | –.29 | –.72 |
| Education and Learning | 7 (5–8) | 8 (6–9) | 14.19*** | 0.39 | 54 | –.35 | –.67 |
| Boundaries and Behaviour | 5 (4–7) | 7 (6–9) | 17.09*** | 0.32 | 62 | .21 | –.51 |
| Family Routine | 7 (5–9) | 8 (7–9) | 13.77*** | 0.24 | 51 | –.45 | –.45 |
| Home and Money | 7 (5–9) | 8 (6–9) | 10.39*** | 0.19 | 42 | –.45 | –.77 |
| Progress to Work | 8 (5–10) | 9 (6–10) | 8.28*** | 0.28 | 33 | –.53 | –1.14 |

Note. IQR = interquartile range. *** $p < .001$.^a Cohen provided rules of thumb for interpreting these effect sizes, suggesting that an r of |.1| represents a small effect size, |.3| represents a medium effect size, and |.5| represents a large effect size.

Table 3. Responsiveness of the Family Star Plus for service Users With the Greatest Need (Beginning With Readings of 1–6)

| Scale | Time 1 Mdn (IQR) | Time 2 Mdn(IQR) | Wilcoxon Z | Effect size ¹ r | Readings improved(%) | N | Skew | Kurtosis |
|----------------------------|------------------|-----------------|------------|----------------------------|----------------------|-----|-------|----------|
| Physical Health | 5 (4–6) | 7 (5–8) | 12.47*** | 0.50 | 75 | 305 | –1.05 | –1.44 |
| Your Well-Being | 5 (3–5) | 7 (5–8) | 16.80*** | 0.50 | 76 | 560 | –.58 | –.67 |
| Meeting Emotional Needs | 5 (4–6) | 7 (5–8) | 14.76*** | 0.49 | 77 | 455 | –.97 | .33 |
| Keeping Your Children Safe | 5 (4–6) | 7 (5–8) | 11.55*** | 0.48 | 77 | 284 | –1.02 | .40 |
| Social Networks | 5 (3–6) | 6 (5–8) | 15.05*** | 0.51 | 75 | 440 | –.66 | –.48 |
| Education and Learning | 5 (3–6) | 7 (5–8) | 14.07*** | 0.48 | 75 | 431 | –.61 | –.61 |
| Boundaries and Behaviour | 5 (4–5) | 7 (5–8) | 17.12*** | 0.49 | 76 | 608 | –.70 | –.05 |
| Family Routine | 5 (4–6) | 7 (5–8) | 14.05*** | 0.50 | 75 | 390 | –.97 | .19 |
| Home and Money | 5 (3–5) | 6 (5–8) | 13.09*** | 0.47 | 71 | 384 | –.75 | –.37 |
| Progress to Work | 4 (2–5) | 5 (3–7) | 11.38*** | 0.41 | 58 | 392 | –.24 | –1.15 |

Note. IQR = interquartile range. *** $p < .001$.

my children’s learning or aspirations: your children aren’t regularly attending school, training or work”) to Effective parenting (“My children are learning well and developing positive aspirations for the future: Their school attendance is good”). The Boundaries and Behaviour area is about parents giving their children clear boundaries and being a positive role model, with stages ranging from Stuck: (“I don’t feel there is a problem with boundaries or behaviour although I know others are concerned”) to Effective Parenting (“I am able to maintain effective boundaries and deal effectively with problems”).

All parents were using a “Tier 2” early help service for families in which children have additional needs and may require extra support

to prevent their needs from becoming more complex or acute. First Stars and absence data were recorded between August 2014 and February 2016. Data were selected for inclusion when percentage absence was recorded after the first Star was completed, resulting a final sample of 263 families. The 125 excluded families had completed Stars before the percentage absence being recorded. In 95% of cases, there were at least 2 months between the Family Star Plus being completed and absence being recorded ($M = 280$ days, $SD = 217.96$, $Mdn = 194$, interquartile range = 101–420). Skewness (.99) and kurtosis (–.06) statistics for the time between these measurements were within the accepted range.

Analytic strategy. Linear regression analysis using SPSS 24.0 was used to examine the associations between Education and Learning and Boundaries and Behaviour and the dependent measure: system recorded unauthorized absence. The regression models were run separately for each of the two predictors, and it was hypothesized that higher readings on the Star in these areas would predict lower unauthorized absence recorded subsequently.

RESULTS

As predicted, there was a significant negative association between Education and Learning readings at the first Star reading and the percentage of school absence, with fewer unauthorized absences among those whose parents had higher readings on the Education and Learning area, $B = -2.24$, $t(261) = -4.49$, $p < .001$. The value of this unstandardized beta coefficient means that on average, every point higher on this Star outcome area was associated with 2.2% fewer unauthorized absences. There also was a significant negative association between Boundaries and Behaviour readings at the first Star reading and the percentage absence recorded, with fewer unauthorized absences among those with higher Boundaries and Behaviour readings, $B = -1.87$, $t(261) = -2.87$, $p = .004$. Therefore, on average, every point higher on the Boundaries and Behaviour area was associated with 1.9% fewer unauthorized absences.

DISCUSSION

This study evaluated interrater reliability, factor structure, internal consistency, item redundancy, responsiveness, and predictive validity of the Family Star Plus. Following from the interrater reliability findings presented by Mackeith (2014), we were able to demonstrate that caseworkers can apply readings using descriptors of Journey of Change stages provided in the Family Star Plus User Guide accurately, with the vast majority showing a high level of chance-corrected reliability. We compared caseworkers readings to those agreed by expert raters, as well as looking at consistency between them because the latter method could produce high interrater reliability when different workers apply the tool consistently but incorrectly. In contrast to the test of interrater reliability for the Recovery Star reported by Killaspy et al. (2012),

in which professionals independently rated service users known to them, we were able also to ensure that the information on which readings were based was standardized and complete enough to enable accurate assessment of the service user described.

The Family Star Plus also was shown to represent a single underlying construct, conceptualised as family functioning, with coherence between the 10 outcome areas demonstrated by high internal consistency and intercorrelations, suggesting no item redundancy. Change in Family Star Plus readings over the course of engagement with services also suggested that the tool is responsive to change, especially among those with the greatest level of need, which is important both from the perspective of measurement and in providing encouragement for service users and providers. Indeed, feedback from the development process of different versions of the Outcomes Star frequently highlights the motivational benefits and the Family Star has been described as ‘helping service users to develop resilience and workers to reflect on their practice in terms of improving outcomes’ (York Consulting, 2013, p. 6).

The possibility that caseworkers may be tempted to inflate change cannot be ignored, especially in the context of payment by results. However, the detailed descriptors in the Outcomes Stars are designed to address this: Caseworkers are trained to refer closely to the detailed descriptions for each point on the scales and to come to agreement on ratings with service users before plotting them on the Star diagram. The association between Star readings and an objective indicator—unauthorized school absence—also suggests that readings are meaningful. Furthermore, research should examine how change on the Star relates to change in other outcomes.

These initial findings linking Star readings to objective outcomes are also of interest because funders often want to see evidence of these outcomes, but organizations find it difficult to access and collate good-quality, objective data. For example, it is an obligation of organizations in receipt of funding from Big Lottery (the National Lottery in the United Kingdom) to demonstrate how they are “making a measurable short-term difference and contributing to making a long-term difference to the lives of people they seek to help.” (Buxton, 2002, p. 2). Although it has been argued that soft outcomes are no less

“real” and have a positive impact on people’s development (e.g., Welsh Council for Voluntary Action, 2002), it is valuable to be able to provide evidence that soft outcomes are in fact associated with longer term impact and outcomes, such as school attendance, employment, and offending. Additional datasets are needed to establish relationships between such outcomes and other areas of the Family Star Plus, but the results presented are encouraging.

CONCLUSIONS

It has been said that “proper measurement requires ... an appropriate methodology for evidence gathering which strikes the right balance between rigour and practicality” (National Youth Agency, 2018, p. 18). In contrast to many outcome measures, completing the Family Star Plus is an integral part of working with service users and is intended to support as well as measure progress during engagement with a service.

This integration of outcomes measurement and casework is important to organizational and individual buy-in and to moving away from measuring outcomes as a “tick-box exercise” and means of surveillance under the norms of the “audit society” (Power, 1999). The Family Star Plus has proved popular in a context in which the measurement of outcomes is increasingly viewed as an essential ingredient in service delivery. Given the wide use of the Family Star Plus, this investigation of its psychometric properties is important in providing evidence that it is a robust outcomes measurement tool.

AUTHOR NOTE

The authors work at Triangle, the social enterprise responsible for creating of the Outcomes Star suite of tools. Joy MacKeith is the Strategic Director and was involved in developing the Family Star Plus. Dr Anna Good joined Triangle as a Research Analyst after the Family Star Plus was developed. The organisations who took part in the case study task and provided Star data pay a license and training fee to Triangle to use the Family Star Plus.

REFERENCES

- Arbuckle, J. L. (2014). Amos (ver. 23.0) [Software program]. IBM SPSS.
- Arvidson, M., & Kara, H. (2013). *Putting evaluations to use: From measuring to endorsing social value* [Working paper]. Third Sector Research Centre. [http://www.birmingham.ac.uk/generic/tsrc/](http://www.birmingham.ac.uk/generic/tsrc/documents/tsrc/working-papers/working-paper-110.pdf)
- Bailey, D., & Kerlin, L. (2015). Can health trainers make a difference with difficult-to-engage clients? A multisite case study. *Health Promotion Practice, 16*(5), 756–764. <https://doi.org/10.1177/1524839915572802>
- Bartlett, M. S., & Jackson, W. (1954). Communication Theory. *Journal of the Royal Statistical Society. Series A (General), 117*(1), 104. <https://doi.org/10.2307/2342774>
- Bradshaw, J., Hoelscher, P., & Richardson, D. (2007a). An index of child well-being in the European Union. *Social Indicators Research, 80*(1), 133–177. <https://doi.org/10.1007/s11205-006-9024-z>
- Bradshaw, J., Hoelscher, P., & Richardson, D. (2007b). Comparing Child Well-Being in OECD Countries: Concepts and Methods (Innocenti Working Paper No. 2006-03). UNICEF Innocenti Research Centre
- Burns, S., & MacKeith, J. (2013). *The Family Star user guide and The Family Star organisation guide*. Triangle Consulting.
- Burns, S., & MacKeith, J. (2017). *The Family Star Plus ser guide and The Family Star Plus organisation guide*. Triangle Consulting.
- Buxton, R. (2002). *The investor approach: A way forward for the Community Fund?* (Introduction). Community Fund.
- Collishaw, S., Maughan, B., Goodman, R., & Pickles, A. (2004). Time trends in adolescent mental health. *Journal of Child Psychology and psychiatry, 45*(8), 1350–1362. <https://doi.org/10.1111/j.1469-7610.2004.00335.x>
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. <https://doi.org/10.1007/bf02310555>
- Department of Communities and Local Government (2012, October). *The Troubled Families programme. Financial framework for the Troubled Families programme’s payment-by results scheme for local authorities*. <https://www.gov.uk/government/publications/financial-framework-for-the-troubled-families-programme-january-2018-onwards>
- Department of Communities and Local Government. (2016, October). *National evaluation of the Troubled Families Programme: Process evaluation*. <https://www.gov.uk/government/publications/national-evaluation-of-the-first-troubled-families-programme>
- Dickens, G., Weleminsky, J., Onifade, Y., & Sugarman, P. (2012). Recovery Star: Validating user recovery. *The Psychiatrist, 36*(2), 45–50. <https://doi.org/10.1192/pb.bp.111.034264>
- Early Childhood Outcomes Center. (2005). *Family and child outcomes for early intervention and early childhood special education*. <http://www.tats>

- .ucf.edu/docs/ECO_Family_Outcomes_Issues_01172005.pdf
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. <https://doi.org/10.1037/1082-989x.4.3.272>
- Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods, 18*(4), 454–474. <https://doi.org/10.1037/a0030005>
- George, D., & Mallery, M. (2010). *SPSS for Windows step by step: A simple guide and reference* (17.0 update; 10th ed.). Pearson.
- Griffiths, C. A., Heinkel, S., & Dock, B. (2015). Enhancing recovery: transition intervention service for return to the community following exit from an alternative to psychiatric inpatient admission—a residential recovery house. *The Journal of Mental Health Training, Education and Practice, 10*(1), 39–50. <https://doi.org/10.1108/jmhtep-09-2014-0027>
- Harris, L., & Andrews, A. (2013). *Implementing the Outcomes Star well in a multi-disciplinary environment*. RMIT University, The Salvation Army, Crisis Services Network.
- Hayes, A.F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1*(1), 77–89. <https://doi.org/10.1080/19312450709336664>
- Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191–205. <https://doi.org/10.1177/1094428104263675>
- Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research. *Educational and Psychological Measurement, 66*(3), 393–416. <https://doi.org/10.1177/0013164405282485>
- Hoffman, L., Marquis, J., Poston, D., Summers, J. A., & Turnbull, A. (2006). Assessing family outcomes: psychometric evaluation of the Beach Center Family Quality of Life Scale. *Journal of Marriage and Family, 68* (4), 1069–1083. <https://doi.org/10.1111/j.1741-3737.2006.00314.x>
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424–453. <https://doi.org/10.1037/1082-989x.3.4.424>
- Kaiser, H. F. (1970). A second generation little jiffy. *Psychometrika, 35*(4), 401–415. <https://doi.org/10.1007/bf02291817>
- Killaspy, H., White, S., Taylor, T. L., & King, M. (2012). Psychometric properties of the Mental Health Recovery Star. *The British Journal of Psychiatry, 201*(1), 65–70. <https://doi.org/10.1192/bjp.bp.111.107946>
- Klett-Davies, M., Skaliotis, E., & Wollny, I. (2009). *Mapping and analysis of parenting services in England: Assessing needs and patterns of spending*. Family and Parenting Institute.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Sage Publications.
- Larsen, J., & Griffiths, C. (2013). Supporting recovery in a third sector alternative to psychiatric hospital admission: Evaluation of routinely collected outcome data. *The Journal of Mental Health Training, Education and Practice, 8*, 116–125. <https://doi.org/10.1108/JMHTEP-04-2013-0016>
- Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, 38*(1), 88–91. <https://doi.org/10.3758/bf03192753>
- MacKeith, J. (2011). The development of the Outcomes Star: A participatory approach to assessment and outcome measurement. *Housing, Care and Support, 14*(3), 98–106. <https://doi.org/10.1108/14608791111199778>
- MacKeith, J. (2014). Assessing the reliability of the Outcomes Star in research and practice. *Housing, Care and Support, 17*(4), 188–197. <https://doi.org/10.1108/hcs-11-2014-0027>
- National Youth Agency. (2018). *The future for outcomes A practical guide to measuring outcomes for young people*. <http://www.nya.org.uk/wp-content/uploads/2014/06/Future-for-Outcomes-Guide1.pdf>
- Nijhof, W. J., & Streumer, J. N. (Eds.). (2012). *Key qualifications in work and education*. Springer Science + Business Media.
- Ógáin, E., Lumley, T. and Prichard, D., 2012. *Making an impact: Impact measurement among charities and social enterprises in the UK*. New Philanthropy Capital.
- Placentino, A., Lucchi, F., Scarsato, G., & Fazzari, G. (2017). Mental Health Recovery Star: Features and validation study of the Italian version. *Rivista di psichiatria, 52*(6), 247–254. <https://doi.org/10.1708/2846.28728>
- Power, M. (1999). *The audit society. Rituals of verification*. Oxford University Press
- Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological assessment, 24*(2), 282–292. <https://doi.org/10.1037/a0025697>
- Schermelleh-Engel, K., Kerwer, M., & Klein, A. G. (2014). Evaluation of model fit in nonlinear multilevel structural equation modeling. *Frontiers in*

- Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.00181>
- Scott, S., Knapp, M., Henderson, J., & Maughan, B. (2001). Financial cost of social exclusion: follow up study of antisocial children into adulthood. *BMJ*, 323(7306), 191. <https://doi.org/10.1136/bmj.323.7306.191>
- Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with parallel analysis. *Psychological Methods*, 16(2), 209–220. <https://doi.org/10.1037/a0023353>
- Welsh Council for Voluntary Action. (2002). *Soft outcomes and European-funded projects*. <https://www.wcva.org.uk/training-and-events/training-latest/2018/06/3sets-measuring-soft-outcomes>
- York Consulting. (2013). *Family Star Evaluation: Summary report*. Family Action.
- Yuan, K.-H., Marshall, L. L., & Bentler, P. M. (2002). A unified approach to exploratory factor analysis with missing data, nonnormal data, and in the presence of outliers. *Psychometrika*, 67(1), 95–121. <https://doi.org/10.1007/bf02294711>