Cross-cultural validity and reliability of the BRIEF-P at age 2 and 4.5 years in children born at risk of neonatal hypoglycaemia

Ryan Jim San Diego[1,3]

Nike Franke[2]

Jane Elizabeth Harding[2]

Trecia Ann Wouldes[1]

for the CHYLD Study Group

[1] Department of Psychological Medicine, The University of Auckland

[2] Liggins Institute, The University of Auckland

[3] Social and Community Health, The University of Auckland

Correspondence should be addressed to: Trecia Wouldes, Department of Psychological Medicine, Faculty of Medical and Health Sciences, The University of Auckland, Private Bag 92019, Auckland 0630, New Zealand.

Email: t.wouldes@auckland.ac.nz

Appendix S1: Further acknowledgements (Not for review)
We wish to acknowledge the contribution of The Children with Hypoglycaemia and their Later Development (CHYLD) Study team: Judith Ansell, Coila Bevan, Nataliia Burakevych, Ellen Campbell, Jessica Charlton, Tineke Crawford, Kelly Fredell, Karen Frost, Claire Hahnhaussen, Safayet Hossin, Greg Gamble, Anna Gsell, Yannan Jiang, Kelly Jones, Sapphire Martin, Chris McKinlay, Grace McKnight, Christina McQuoid, Chris McKinlay, Neil Mickelwood, Janine Paynter, Jenny Rogers, Ryan San Diego, Kate Sommers, Heather Stewart, Anna Timmings, Jess Wilson, from the Liggins Institute, University of Auckland; Nicola Anstice, Arijit Chakraborty, Robert Jacobs, Gill Matheson, Nabin Paudel, Sandy Yu, from the Department of Optometry and Vision Science, University of Auckland. Max Berry, Arun Nair, Alexandra Wallace, Phil Weston from the Department of Paediatrics, Waikato Hospital, Hamilton, New Zealand. Aaron Le Compte, Department of Engineering, University of Canterbury. International Advisory Group: Heidi Feldman, Stanford University School of Medicine, USA; William Hay, University of Colorado School of Medicine, USA; Darrell Wilson, Stanford University School of Medicine, USA; Robert Hess, McGill Vision Research Unit, Department of Ophthalmology, McGill University, USA.

ABSTRACT

Executive function (EF) encompasses several neurocognitive processes that are important in self-regulation of behaviour and the attainment of social and cognitive competencies. While much progress has been made in developing valid measures of adult and adolescent EF, there is a dearth of valid measures for preschool children. Given the steep trajectory of neuropsychological development among this age group and the importance of EF, a valid measure for clinical assessment and research is needed that can capture EF in the everyday context of early childhood. The Behaviour Rating Inventory of Executive Function Preschool Version (BRIEF-P) measures parent and teacher observations of children's everyday self-regulatory behaviours. The BRIEF-P has been validated in a range of normative and non-normative samples, but further validation is needed across cultures. This study aimed to evaluate the cross-cultural validity and reliability of the BRIEF-P when used by New Zealand Māori (*n* = 131*)* and European (*n* = 193) parents of children born with risk factors of neonatal hypoglycaemia. Parents of children who participated in the prospective, longitudinal Children with Hypoglycaemia and their Later Development (CHYLD) study completed the BRIEF-P when the child was 2 years ± 4 weeks and 4.5 years ± 8 weeks old. Results showed the BRIEF-P is a highly reliable and valid instrument. Comparisons between Māori and New Zealand European samples revealed biases which could be a source of further work to improve the construct validity of this measure such as the development of norms and item validation for non-European, non-Western samples.

Key words: BRIEF-P, cross-cultural, reliability, validity, executive function

INTRODUCTION

Executive function (EF) or executive control is a theoretical construct that incorporates higher order cognitive processes such as the purposeful maintenance of information in memory, inhibition of automatic or prepotent responses, and shift and maintenance of attention to achieve a goal or solve a problem, especially in novel or uncertain situations (Diamond, 2013; Espy & Bull, 2005). Relying on the prefrontal cortex (Jurado & Rosselli, 2007), EF develops rapidly during early childhood and continues to develop well into young adulthood (Diamond, 2016). The preschool years constitute a period in the development of EF where surges in performance on several EF tasks are observed and prefrontal neural systems linked to EF show a gradual differentiation into distinct functional systems (Best & Miller, 2010; Carlson, 2005; Posner, Rothbart, Sheese, & Voelker, 2012; Rubia, 2013). Current literature suggests normal EF development follows a sequence with basic inhibitory and working memory abilities (inhibition of a prepotent response and maintenance of information) appearing during the first year after birth followed by more complex forms of these by age 3 years (Carlson, 2005; Garon, Bryson, & Smith, 2008). Shifting, the ability to shift attention and transition between tasks, is dependent on these early abilities and emerges around age 3 years followed by early abilities in planning and organisation that have been observed between ages 3 and 4 years (Espy, Kaufmann, Glisky, & McDiarmid, 2001; Zelazo et al., 2003). The regulation of emotional responses is thought to develop together with other EF abilities and steadily improve across childhood (Posner et al., 2012).

EF plays an important role in cognitive, social, and emotional development (Blair & Razza, 2007; Gioia, Isquith, & Kenealy, 2010) and is thought to be a better predictor of school readiness than intelligence (Blair & Razza, 2007). Longitudinal

research has shown that EF in early childhood predicts long-term health, educational outcomes, and wealth (Diamond, 2016; McClelland, Acock, Piccinin, Rhea, & Stallings, 2013), whereas, deficits in EF can lead to problematic outcomes such as criminal offending, substance dependence, and lower socioeconomic status (Diamond, Barnett, Thomas, & Munro, 2007; Moffitt et al., 2011). In addition, EF deficits are associated with a wide range of psychopathologies such as attention deficit hyperactivity disorder (Kasper, Alderson, & Hudec, 2012), conduct disorder (Morgan & Lilienfeld, 2000), depression (Snyder, 2013), and schizophrenia (Mesholam-Gately, Giuliano, Goff, Faraone, & Seidman, 2009).

However, EF is open to remediation through interventions that provide opportunities to practice these skills, particularly during early childhood (Bierman & Torres, 2016; Diamond et al., 2007; Scionti, Cavallero, Zogmaister, & Marzocchi, 2020). In order to identify the deficits in EF and the effects of remediation in early childhood, sensitive and reliable measures of early EF are needed that can capture the variable nature of behaviour in different contexts, differences in motor and verbal ability during this period of rapid development (Carlson, 2005), and the differential rate of development of each EF skill (Blackwell, Chatham, Wiseheart, & Munakata, 2014; Friedman & Miyake, 2017). Additionally, it is important that EF measures are ecologically and cross-culturally valid. A measure is ecologically valid when it reliably measures the same constructs in real world settings or day-to-day life as it does in research settings (Gioia, Espy, & Isquith, 2003), and cross-culturally valid when it reliably measures the same constructs across different cultural contexts (Hughes, 1990).

The Behavior Rating Inventory of Executive Function-Preschool Version (BRIEF-P) is a parent, teacher or other frequent caregiver report that measures

specific components of EF through observable everyday behaviours in children 2 to 5 years 11 months. It is increasingly being used to assess EF in preschool aged children (Gioia et al., 2003; Gioia & Isquith, 1996). The BRIEF-P has five clinical scales (Inhibit, Shift, Emotional Control, Working Memory, Plan/Organize), three composite indices (Inhibitory Self-Control Index (ISCI), Flexibility Index (FI), Emergent Metacognition Index (EMI)), and an overall composite score known as the Global Executive Composite (GEC).

The BRIEF-P is used extensively in clinical settings within multiple subgroups of children, but its psychometric properties have mainly been tested in normally developing children in North America and Europe. Only one validation study has involved a clinical sample of children (Bausela, 2019), and cross-cultural validity has only been tested in two studies (Ebrahimi, Abedi, Yarmohammadian, & Faramarzi, 2016; Kown, 2017). In one, an Iranian study, internal consistency was satisfactory and acceptable model fit was achieved after removing two items (Ebrahimi et al., 2016). The other was published in Korean and showed a good internal consistency and model fit with the five subscales (Kown, 2017). Finally, to date all evidence for the validity of the BRIEF-P comes from cross-sectional studies that test validity in populations at one age or pooled data from children aged 3-6. Given the significant changes in all cognitive processes and the surge in the development of EF over this period it is likely that the structural organisation of EF changes during this developmental stage as well.

North American and European validation studies have reported mixed findings for the psychometric properties of the BRIEF-P. Good internal consistency for the scales have been reported in a number of studies (Bausela, 2019; Duku & Vaillancourt, 2014; Gioia et al., 2003; Spiegel, Lonigan, & Phillips, 2017).

Convergent validity has also been reported with the Child Behavior Checklist (CBCL) subscales (Duku & Vaillancourt, 2014), with academic ability (Test of Preschool Early Literacy and Bracken school Readiness Assessment), and with performance-based measures of EF (Head Toes Knees Shoulders) (Spiegel et al., 2017). However, findings on the construct validity of the BRIEF-P are inconsistent. There is evidence for the existence of a one-order five-factor model together with a second-order three-factor solution as identified by the authors of the BRIEF-P (Bonillo, Jimenez, Ballabriga, Capdevila, & Riera, 2012; Ezpeleta, Granero, Penelo, de la Osa, & Domènech, 2015; Gioia et al., 2003), with the three-factor model confirmed in a Norwegian sample of 1134 3-year-old children. However, the Norwegian study identified an almost equally suitable second-order one-factor model (Skogan et al., 2016). In another study a series of confirmatory factor analyses (CFA) revealed a four-level structure and a bi-factorial model as best solutions, indicating the items did not map onto factors as expected (Spiegel et al., 2017). A Canadian replication study revealed that three of the five clinical scales (i.e., Emotional Control, Plan/Organize, and Working Memory) were unidimensional and that the two remaining scales were multidimensional, thus failing to confirm the factor structure found by the BRIEF-P authors (Gioia et al., 2003).

Current study

This study aimed to evaluate the cross-cultural validity and reliability of the BRIEF-P when used by New Zealand (NZ) European and Māori parents of preschool-aged children at 2 and 4.5 years of age enrolled in the Children with Hypoglycaemia and their Later Development (CHYLD) study. All children enrolled in this study were born with risk factors for neonatal hypoglycaemia. Results at ages 2 and 4.5 found no differences between children who met criteria for hypoglycaemia and those who did

not in the percent of children with *T* scores greater than 65 on the BRIEF-P (McKinlay et al., 2015; McKinlay et al., 2017). However, a lower composite score on five examiner administered tests of EF was associated with severe hypoglycaemia . Using BRIEF-P data from the CHYLD study, we examined: 1) the cross-cultural validity of the BRIEF-P by comparing the scores of NZ European and Māori on the clinical scales at both time points and the associations between the BRIEF-P clinical scales across time points by ethnicity; 2) the reliability (internal consistency and test-retest reliability) of the BRIEF-P by ethnicity at two time points; 3) whether a second-order five-factor model proposed by the authors of the BRIEF-P fit the data similarly across ethnicities and at both time points, and; 4) the convergent and incremental validity of the BRIEF-P at age 2 and 4.5 years with the Child Behavior Checklist (CBCL) at 4.5 years for each ethnic group.

METHOD

Participants

Participants were 324 children of NZ European (*n* = 193) or Māori (*n* = 131) descent enrolled in the CHYLD study. This prospective, longitudinal study included children born with one or more known risk factors for neonatal hypoglycaemia: maternal diabetes, preterm (<37 weeks' gestation), small birth weight (<10th percentile or <2500 g), and large birthweight (>90 percentile or >4500 g). Children in the CHYLD cohort were born at Waikato Hospital, Hamilton, New Zealand between 2006 and 2010. Further details are published elsewhere (Burakevych et al., 2017; McKinlay et al., 2017).

--- TABLE 1 ---

Procedure

Ethics approval was obtained from the Northern Y Regional Health and Disability Ethics Committee (NTY/10/03/021). Written consent was obtained at study entry and at each follow-up. Parents completed questionnaires when the children were 2 years (± 4 weeks) and 4.5 years' (± 8 weeks) corrected (postmenstrual) age.

Measures

A demographics questionnaire was completed by children's primary caregiver. Home address was used to obtain families' New Zealand Deprivation Index, which includes access to internet, income, receiving a government benefit, employment status, education, home ownership, single parent family, and adequacy of housing using census data to yield a socioeconomic decile ranging from 1 to 10, with higher scores indicative of higher levels of deprivation.

The *Behavior Rating Scale of Executive Function – Preschool version* (BRIEF-P; Gioia et al., 2003) was used to measure everyday EF at both time points. Parents reported how often a specific behaviour had occurred over the past 6 months. The BRIEF-P consists of 63 items rated on a 3-point Likert scale (never; sometimes; often) and has five clinical scales: *Inhibit* (16 items) measures the ability to control prepotent behavioural responses in everyday situations; *Shift* (10 items) measures the ability to switch from a situation or task to another; *Emotional Control* (10 items) provides an indication of EF within the emotional domain and measures emotional reactions in varying situations; *Working Memory* (17 items) measures the capacity to retain information in order to solve a problem or complete a task that requires a number of steps; *Plan/Organize* (10 items) measures the ability to *Plan* or look forward to future events and carry out instructions to meet goals and figure out the steps needed to complete a task, *Organize* relates to the ability to organise information, actions or materials to reach a goal. These 5 scales can be summarised

into 3 indices: *Inhibitory Self-Control, Flexibility,* and *Emergent Metacognition* and a total score termed *Global Executive Composite* (GEC). Higher scores on any of the clinical scales indicate poorer EF. A *T* score of 65 and over is considered clinically significant. The extent to which respondents answer questions inconsistently, as compared to combined community and clinical samples, is measured by the Inconsistency scale. The Negativity scale detects unusual negative answers of a selection of BRIEF-P items.

The *Child Behavior Checklist 1.5-5* (Achenbach & Rescorla, 2000) is a 99-item measure that evaluates child internalising and externalising problems within the past 2 months. The CBCL has the following seven subscales: *Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems*, and *Aggressive Behavior*. In the current study internal consistencies were moderate to high (r = .68 to .92). Total scores of 70 and over are considered within the clinical range. Parents completed the CBCL at 4.5 years only.

Analytic Plan

Statistical analyses were performed using IBM SPSS version 27 and JASP version 0.14.1. To test cross-cultural validity a series of *t*-tests were used to compare scores between NZ European and Māori on the BRIEF-P clinical scales (Inhibit, Shift, Emotional Control, Working Memory and Plan/Organize), the summary indices of Inhibitory Self-control, Flexibility, Emergent Metacognition and the total score, the Global Executive Composite at age 2 and 4.5 years. Effect sizes were expressed in Cohen's *d*. Test-retest reliabilities were obtained by examining the relationships within and between the BRIEF-P clinical scales at both time points for each ethnic group using Pearson correlation coefficients. Internal consistency of the BRIEF-P

clinical scales across both time points and ethnicities was calculated by Cronbach's alpha coefficient and McDonald's Omega (ω). Multi-group confirmatory factor analyses (MGCFA) were employed to evaluate the structural invariance of the BRIEF-P at each time point, using ethnicity as the grouping variable. We examined whether a second-order five-factor model would fit the data similarly across ethnicities. We used multiple indicators of model fit, including the Comparative Fit Index (CFI ≥ 0.95), Tucker Lewis Index (TLI ≥ 0.95, Root Mean Square Error of Approximation (RMSEA ≤ 0.06), Standardised Root Mean Residual (SRMR ≤ 0.08), and the likelihood ratio chi-square ($\chi 2/df \leq 3$) (Hu & Bentler, 1999). In testing for equivalence with MGCFA four steps were considered: First, configural invariance or equivalence of model form; second, metric invariance or equivalence of factor loadings; third, scalar invariance or equivalence of item intercepts, and; fourth, strict invariance or equivalence in item residuals (Fischer & Karl, 2019; Putnick & Bornstein, 2016). Convergent and incremental validity of the BRIEF-P with the CBCL was examined for each ethnicity. To test convergent validity, we calculated bivariate correlations between BRIEF-P clinical scales at 2 and 4.5 years and the CBCL subscales at 4.5 years. To test incremental validity, we conducted a series of hierarchical multiple regressions for each CBCL subscale, with BRIEF-P at age 2 years entered at step 1 and BRIEF-P at age 4.5 years entered at step 2.

RESULTS

In comparison to the NZ European group, NZ Māori families had a lower level of maternal education and were more likely to receive financial support from the government and reside in a more deprived neighbourhood (Table 1).

Cross-cultural validity

As shown in Table 2, there were significant differences between ethnicities at age 2 years, with Māori parents reporting higher scores than NZ European parents on the overall total score, GEC, $t(317) = 2.13$, $p = .034$, the Inhibitory Self-Control Index, $t(317) = 2.20$, $p = .024$, and the Inhibition clinical scale, $t(317) = 2.51$, $p = .013$. At age 4.5 years, Māori parents also reported a higher overall total score than NZ European parents, GEC, $t(317) = 3.45$, $p = .001$, and higher scores on the Inhibitory Self-Control, $t(317) = 2.58$, $p = .010$, Flexibility, $t(317) = 2.36$, $p = .019$, and, Emergent Metacognition, $t(317) = 3.84$, $p < .001$ indices and on the Inhibit, $t(317) = 2.66$, $p = .008$, Shift, $t(317) = 2.54$, $p = .012$, Working Memory, $t(317) = 4.26$, $p < .001$, and Plan/Organize, $t(317) = 2.60$, $p = .010$ clinical scales. These results suggest Māori parents' rated their children as having poorer everyday EF skills than NZ European parents, however, these findings could also reflect other social determinants that have historically been found to differ between these groups. No significant differences between ethnicities were found in the rate of negativity and inconsistency scores at either time point. At age 2 years there were 4.3% cases of negativity and 4.0% inconsistency and at age 4.5 years these percentages were 2.8% and 4.6%, respectively.

--- TABLE 2 ---

Reliability of the BRIEF-P clinical scales at 2 and 4.5 years

Test-retest reliabilities were computed to identify whether clinical scores on the BRIEF-P were stable between 2 and 4.5 years of age and to determine whether developmental changes were present. Intra-scale positive correlations within the NZ European and Māori groups were significant and ranged from weak to moderate ($r = .36$ to $.49$) across time (Table 3). For Māori, moderate intra-scale correlations from 2 to 4.5 years were observed for three of the clinical scales, Inhibition ($r = .43$),

Emotional Control (r = .49), and Plan/Organize (r = .41). Moderate intra-scale correlations for NZ European were found for all five of the clinical scales, Inhibition (r = .48), Shift (r = .50), Emotional Control (r = .40), Working Memory (r = .50), and Plan/Organize (r = .43).

As EF skills advance rapidly during early childhood, we also examined whether inter-scale correlations would be stronger at 4.5 years compared to 2 years. Inter-scale positive correlations were weak to moderate for both ethnicities at age 2 years. At age 4.5 years, moderate to strong positive inter-scale correlations were found for both ethnicities, with the highest correlations reported for Inhibition and Emotional Control (r = .73, Māori; r = .66, NZ European), Inhibition and Working Memory (r = .72, Māori; r = .72, NZ European), Inhibition and Plan/Organize (r = .77, Māori; r = .62, NZ European), and Working Memory and Plan/Organize (r = .83, Māori; r = .78, NZ European). For Māori, strong correlations were observed between Emotional control and Working Memory at age 4.5 years (r = .64) and Plan/Organize (r = .69), while these correlations were only moderate for NZ European. The substantial increase in correlations observed for both ethnic groups at 4.5 years shows that the test-retest reliability of the BRIEF-P is good even after a duration of two years.

The internal consistency of the BRIEF-P is also reported in Table 3. Cronbach's alphas for the BRIEF-P clinical scales ranged from good to very good for the Māori group (α = .75 to .87 at age 2 years; α = .82 to .89 at age 4.5 years) and from good to excellent for the NZ European group (α = .78 to .87 at age 2 years; α = .83 to .96 at age 4.5 years). Overall scale consistency ranged from α = .94 to .96 across ethnicities and timepoints.

Omega coefficients from the factorial framework were weak to excellent for both the Māori group (ω = 0.29 to 0.79 at age 2 years, ω = 0.67 to 0.94 at age 4.5 years) and the NZ European group (ω = 0.22 to 0.99 at age 2 years, ω = 0.37 to 0.88 at age 4.5 years). The omega coefficients estimated the covariance of items being accounted for in a one-factor model. These results showed that some clinical scales had consistently low reliabilities across ethnic groups and at both time points, namely Shift and Emotional Control, and consistently strong reliabilities for Working Memory and Plan/Organize. These findings may have implications for further measurement analysis for BRIEF-P, such that constructs pertaining to Shift and Emotional Control may have sparse item-to-construct associations.

Construct validity

We evaluated a baseline, second-order 5-factor model (5 clinical scales and 1 global index) using multi-group confirmatory factor analyses (MGCFA) with ethnicity as the grouping variable at 2 and 4.5 years. Both NZ European and Māori results showed poor to fair fit (NZ European, χ^2/df = 2.87, CFI = 0.69, RMSEA = 0.06, SRMR = 0.08 at 2 years; χ^2/df = 4.80, CFI = 0.76, RMSEA = 0.06, SRMR = 0.08 at 4.5 years; and Māori, χ^2/df = 2.55, CFI = 0.57, RMSEA = 0.08, SRMR = 0.09 at 2 years; χ^2 = 3.52, CFI = 0.69, RMSEA = 0.07, SRMR = 0.08 at 4.5 years). In order to test the invariance of this model, we first tested the qualitative feature of the constructs of the BRIEF-P (configural invariance) and found measurement non-invariance across time points (RMSEA = 0.07, CFI = 0.63 at 2 years; RMSEA = 0.06, CFI = 0.74 at 4.5 years). Therefore, we could not proceed to test scalar, metric or strict invariance (Fischer & Karl, 2019). Given these findings, the construct validity of the BRIEF-P for this sample could not be fully ascertained against the second-order five-factor model likely due to differences between ethnic groups.

--- Table 4 (CFA invariance) ---

Convergent validity

Correlations between the BRIEF-P clinical scales at ages 2 and 4.5 and the CBCL clinical scales were calculated to examine convergent validity (Table 5). Overall, stronger correlations were observed between BRIEF-P and CBCL clinical scales at 4.5 years for both NZ European and Māori groups. For instance, in the Māori group strong associations were found between Inhibit, Emotional Control, and Plan/Organize on the BRIEF-P and Aggressive Behavior (r = 0.71, r = 0.72, and r = 0.69, respectively) (r = 0.71) on the CBCL, Inhibit and Attention Problems (r = .68), and Emotional Control on the BRIEF-P and Emotionally Reactive on the CBCL (r =.66). A similar pattern was observed with the NZ European group. Emotional Control on the BRIEF-P was associated with Aggressive Behavior (r = .73) and Emotionally Reactive (r = .69) on the CBCL. Inhibit also had a strong correlation with Attention Problems (r = .68) and Aggressive Behavior (r = .65) on the CBCL, Shift and Emotionally Reactive (r = .66) and Shift with Anxious/Depressed (r = .66) were highly correlated with externalising behaviours.

Group differences were found between Māori and NZ European for somatic complaints. Māori had non-significant positive correlations on the BRIEF-P clinical scales at 2 years with somatic complaints on the CBCL (r = 0.00 to r = 0.15), while the NZ European group had significant positive, but weak to moderate correlations (r = 0.24 to r = 0.37) with somatic complaints.

A series of hierarchical regression analyses were performed to evaluate incremental validity of the BRIEF-P clinical scales, in predicting The CBCL scales Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems and Aggressive Behavior (Supplemental Table A1).

Results revealed that BRIEF-P at 2 years and BRIEF-P at 4.5 years significantly, and independently, contributed to the prediction of the CBCL clinical scales, suggesting that developmental changes of executive function can predict child behaviour problems.

We then conducted hierarchical regressions by ethnicity, with the five clinical BRIEF-P scales at age 2 years entered at step 1 (model 1) and the 4.5-year scales entered at step 2 (model 2). This sequential process allowed us to provide evidence that parent reports of their children's executive function at 2 years and at 4.5 years significantly contributed towards prediction of both internalising and externalising behaviours. We found internalising behaviours (Emotionally Reactive, Anxious/Depressed, Withdrawn) and externalising behaviours (Aggressive Behavior and Attention Problems) are highly predicted in each model for both Māori and NZ European groups. This implies that developmental assessments and changes in executive function as observed by parents during the early years are relevant predictors of early childhood behaviour problems, but with cultural differences. For Māori, emotionally reactive, anxious depressed, and aggressive behaviours were significant predictors as early as 2 years and again at 4.5 years of age. For NZ European, prediction of these did not occur until 4.5 years, providing preliminary evidence that Māori parents in our cohort may be more likely to observe emotion modulation and related regulation processes as early as the toddlerhood stage.

Shift was also found to be a significant predictor of internalising problems, while Inhibit, Working Memory and Plan/Organize were predictive of externalising issues at 4.5 years among the Māori group and also among the NZ European group. These findings suggest that shift is an important mechanism of executive function that shows young children's ability to tolerate distress, switch their attention or make

adjustments that are relevant in the development of anxiety and depression. Similarly, the capacity to hold information (rules or social norms), goal-oriented behaviours, and the ability to resist impulses are more in line with the attention problems and hostile/ aggressive behaviours in young children.

DISCUSSION

This is the first study to determine whether the psychometric properties of the BRIEF-P are similar across two cultures and whether these properties change with development during a period where EF abilities are rapidly developing. Cross-cultural validity, test-retest reliability, and construct, convergent and incremental validity of the BRIEF-P were examined using reports from NZ European and Māori parents whose children were born with risk factors associated with neonatal hypoglycaemia.

Cross-cultural validity

Significant differences in EF behaviour were reported by Māori compared to NZ European parents at both 2 and 4.5 years of age. Māori parents reported a higher overall total score (GEC) on the BRIEF-P at 2 and 4.5, and higher scores on the Inhibition clinical scale and the composite index of inhibitory behaviour (Inhibitory Self-Control Index). These findings suggest that Māori children had more difficulty than European children inhibiting prepotent responses in everyday situations. Additionally, at 4.5 years Māori parents compared to European parents rated their children as being less capable on a wider range of EF abilities than at 2 years. Scores on the clinical scales of Shift, Working Memory and Plan/Organize indicated Māori children were rated as more likely than European children to exhibit behaviours that indicated problems with the ability to be flexible and shift from one task to another. Higher scores on Working Memory and Plan/Organize indicated difficulties in the

ability to hold and manipulate information in short term memory in order to solve a problem, and the ability to carry out instructions to meet goals and determine the steps needed to complete a task.

These results may indicate Māori parents perceive their children as exhibiting more behaviours that are associated with less competence in EF abilities at ages 2 and 4.5. However, these results could also reflect the social and economic differences between Māori and European cultures that have been found to occur in New Zealand (Marriott & Alinaghi, 2021).

Reliability of the BRIEF-P clinical scales from ages 2 to 4.5

The internal consistency of the BRIEF-P was very good across ethnicity and time (Table 3). Despite a 2.5-year interval between parent reports, intra-scale correlations of the BRIEF-P clinical scales showed moderate associations for Māori between 2 and 4.5 years on the clinical scales of Inhibition, Emotional control and Plan/Organize. Moderate associations were also observed for the NZ European sample for all five of the clinical scales, Inhibit, Shift, Emotional Control, Working Memory and Plan/Organize.

Our findings of weak to moderate inter-scale correlations between parent reports at ages 2 and 4.5 years suggest that the BRIEF-P captures changes in the development of EF across infancy and early childhood in both NZ European and Māori children. However, inspection of specific inter-scale correlations by ethnicity and time revealed that differences in EF development between cultures can be identified as early as the preschool years. In comparison, Skogan et al. (2016) reported similar correlations in a Norwegian sample of 3-year-olds to our 2-year data and Bausela (2019) found similar results of parent ratings in a sample of 4.5-year-old children compared to our 4.5-year findings.

Construct validity

A test of the second-order, 5- factor model (5 clinical scales and 1 global index) proposed by the test developers found poor to fair fit for both NZ Māori and European samples. However, non-invariance of these measurement models (configural invariance) was found across time points, and therefore we could not test scalar, metric or strict invariance. This means the pattern of item loadings on each clinical scale differs between NZ European and Māori groups. These findings could be a result of a limited sample size for each group resulting in less observations against the number of parameters of this complex model. Nevertheless, these results support earlier findings showing the items of the BRIEF-P do not consistently map onto factors in the way that would be expected based on its purported 5-factor structure (Isquith, Gioia, & Espy, 2004). Bonillo et al. (2012) found their 5-factor model of a translation of the BRIEF-P into Catalan failed to converge and suggested fewer dimensions underlie the BRIEF-P, as did Spiegel et al. (2017) who suggested that the subscales had little utility beyond the measurement of a global self-regulatory capacity. Duku and Vaillancourt (2014) found that neither a five-factor nor a one-factor model provided adequate fit. Ezpeleta et al. (2015) found support for a five-factor model, but only after removing four questions from the BRIEF-P.

Convergent and incremental validity

Convergent validity is generally considered adequate when correlations with an instrument measuring the same construct is >0.50. Weak correlation coefficients at age 2 years across both ethnicities (range $r = 0.00$ to $r = 0.27$) provided little evidence of convergent validity with the CBCL. However, findings from the correlations between BRIEF-P and the CBCL as well as the hierarchical regression analyses at age 4.5 years do provide evidence for both convergent and incremental validity. In both

NZ Māori and European samples, the individual contribution of the BRIEF-P scales in the prediction of separate CBCL scales showed that Shift appears to be the best predictor, followed by Inhibit and Emotional Control. In NZ Māori the BRIEF-P Working Memory and Plan/Organize correlates well with the CBCL Attention Problems, Withdrawn, and Emotionally Reactive. Consistent with Sherman and Brooks (2010) and Duku and Vaillancourt (2014), Working Memory and Plan/Organize had the same associations between Attention Problems and Withdrawn, but the association with Emotionally Reactive was only observed in the Sherman and Brooks report.

Implications

These results have implications for the use of the BRIEF-P in child assessment practices in that measuring executive function skills for both groups at the level of the five clinical scales is more appropriate than using the indices. Using the clinical scales may help in identifying specific skill deficits that are not identifiable when using the global or index scores and may aid in the development of appropriate interventions.

CONCLUSION

We were able to show that the BRIEF-P is a reliable and valid measure of everyday executive function for New Zealand European and Māori children. However, there were cultural differences in parent reports. We showed that the BRIEF-P has a multidimensional construct that is useful in predicting problem behaviour in children and provided evidence that construct bias could result in higher prediction of problem behaviour in NZ Māori children. Therefore, consideration should be given to further examination of the norms and standardisation in Non-European and non-Western samples of children. In addition, some consideration

should be given to the level of measurement (items, clinical scales and construct validity) and at the level of clinical and developmental significance (parents' perception of child development and culture). Addressing these could improve the ecological and cross-cultural validity of the BRIEF-P. Finally, when interpreting the cross-cultural findings of the BRIEF-P it is important to consider social determinants as potentially contributing to parent ratings as there were statistically significant differences between groups for education and other socioeconomic variables.

Acknowledgements

The authors would like to thank the parents who participated in this study and the members of the CHYLD (Children with Hypoglycaemia and their Later Development) Study Team (Appendix S1). We would particularly like to thank members of the Steering Committee who read and made suggestions about the analysis and content of this paper including, Jane Alsweiler, Department of Paediatrics; Child and Youth Health, University of Auckland, Gavin Brown, Faculty of Education and Social Work, University of Auckland. J. Geoffery Chase, Department of Engineering, University of Canterbury, Jane Harding, Liggins Institute, University of Auckland, Deborah Harris, Newborn Intensive Care Unit, Waikato District Health Board, Ben Thompson, School of Optometry and Vision Science, University of Waterloo.

Funding

This work was supported by grant R01HD069622 from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, by grant 10-399 from the Health Research Council of New Zealand, by the Auckland Medical Research

Foundation, and by Gravida, National Research Centre for Growth and Development, New Zealand.

References:

Bausela, E. (2019). BRIEF-P: Validation study in children in early childhood with neurodevelopmental disorders. *SAGE Open, 9*(3), 2158244019879166. doi:0.1177/2158244019879166

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child development, 81*(6), 1641-1660. doi:10.1111/j.1467-8624.2010.01499.x

Bierman, K. L., & Torres, M. (2016). Promoting the development of executive functions through early education and prevention programs. In J. A. Griffin, L. S. Freund, & P. McCardle (Eds.), *Executive function in preschool age children: Integrating measurement, neurodevelopment and translational research* (pp. 299-326). Washington, DC: American Psychological Association.

Blackwell, K. A., Chatham, C. H., Wiseheart, M., & Munakata, Y. (2014). A developmental window into trade-offs in executive function: The case of task switching versus response inhibition in 6-year-olds. *Neuropsychologia, 62*, 356-364. doi:10.1016/j.neuropsychologia.2014.04.016

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child development, 78*(2), 647-663. doi:10.1111/j.1467-8624.2007.01019.x

Bonillo, A., Jimenez, E. A. A., Ballabriga, M. C. J., Capdevila, C., & Riera, R. (2012). Validation of Catalan version of BRIEF-P. *Child neuropsychology, 18*(4), 347-355. doi:10.1080/09297049.2011.613808

Burakevych, N., Mckinlay, C. J. D., Alsweiler, J. M., Wouldes, T. A., Harding, J. E., & Chyld Study Team. (2017). Bayley-III motor scale and neurological examination at 2 years do not predict motor skills at 4.5 years. *Developmental Medicine & Child Neurology, 59*(2), 216-223. doi:216–223. doi:10.1111/dmcn.13232

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental neuropsychology, 28*(2), 595-616. doi:10.1207/s15326942dn2802_3

Diamond, A. (2013). Executive functions. *Annual review of psychology, 64*, 135-168. doi:10.1146/annurev-psych-113011-143750

Diamond, A. (2016). Why improving and assessing executive functions early in life is critical. In J. A. Griffin, P. McCardle, & L. S. Freund (Eds.), *Executive function in preschool age children: Integrating measurement, neurodevelopment and translational research*. doi:10.1037/14797-002

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science, 318*(5855), 1387-1388. doi:10.1126/science.1151148

Duku, E., & Vaillancourt, T. (2014). Validation of the BRIEF-P in a sample of Canadian preschool children. *Child neuropsychology, 20*(3), 358-371. doi:10.1080/09297049.2013.796919

Ebrahimi, A., Abedi, A., Yarmohammadian, A., & Faramarzi, S. (2016). Psychometric properties of the Behavior Rating Inventory for Executive Functions (BRIEF-P) among preschool children. *Dev Pschology Iran Psychol, 12*(48), 427-439. doi:10.17795/intjsh29022

Espy, K. A., & Bull, R. (2005). Inhibitory processes in young children and individual variation in short-term memory. *Developmental neuropsychology, 28*(2), 669-688. doi:10.1207/s15326942dn2802_6

Espy, K. A., Kaufmann, P. M., Glisky, M. L., & McDiarmid, M. D. (2001). New procedures to assess executive functions in preschool children. *The Clinical Neuropsychologist, 15*(1), 46-58. doi:10.1076/clin.15.1.46.1908

Ezpeleta, L., Granero, R., Penelo, E., de la Osa, N., & Domènech, J. M. (2015). Behavior Rating Inventory of Executive Functioning–Preschool (BRIEF-P) applied to teachers: Psychometric properties and usefulness for disruptive disorders in 3-year-old preschoolers. *Journal of Attention Disorders, 19*(6), 476-488. doi:10.1177/1087054712466439

Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology, 10*, 1507. doi:10.3389/fpsyg.2019.01507

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86*, 186-204. doi:10.1016/j.cortex.2016.04.023

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: a review using an integrative framework. *Psychological bulletin, 134*(1), 31. doi:10.1037/0033-2909.134.1.31

Gioia, G. A., Espy, K. A., & Isquith, P. K. (2003). *Behavior rating inventory of executive function, preschool version (BRIEF-P)* Lutz, FL: Psychological Assessment Resources.

Gioia, G. A., & Isquith, P. K. (1996). *Behavior rating inventory of executive function-preschool version (BRIEF-P)*.

Gioia, G. A., Isquith, P. K., & Kenealy, L. E. (2010). Assessment of behavioral aspects of executive function. In V. Anderson, R. Jacobs, & P. Anderson (Eds.), *Executive Functions and the Frontal Lobes: A Lifespan Perspective* (pp. 179). Hove, UK: Psychology Press.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Hughes, C. C. (1990). Rethinking psychiatry: from cultural category to personal experience. *The Journal of Nervous and Mental Disease, 178*(3), 210-211.

Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behavior. *Developmental neuropsychology, 26*(1), 403-422. doi:10.1207/s15326942dn2601_3

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology review, 17*(3), 213-233. doi:10.1007/s11065-007-9040-z

Kasper, L. J., Alderson, R. M., & Hudec, K. L. (2012). Moderators of working memory deficits in children with attention-deficit/hyperactivity disorder (ADHD): A meta-analytic review. *Clinical psychology review, 32*(7), 605-617. doi:10.1016/j.cpr.2012.07.001

Kown, Y.-H. (2017). A preliminary study on validity of Korean Behavior Rating Inventory of Executive Function-Preschool Version (BRIEF-P). *Journal of Fisheries and Marine Sciences Education, 29*(2), 436-446. doi:10.13000/JFMSE.2017.29.2.436

Marriott, L., & Alinaghi, N. (2021). Closing the gaps: An update on indicators of inequality for Māori and Pacific people. *The Journal of New Zealand Studies*, 2-39. doi:10.26686/jnzs.iNS32.6863

McClelland, M. M., Acock, A. C., Piccinin, A., Rhea, S. A., & Stallings, M. C. (2013). Relations between preschool attention span-persistence and age 25

educational outcomes. *Early Childhood Research Quarterly, 28*(2), 314-324. doi:10.1016/j.ecresq.2012.07.008

McKinlay, C. J., Alsweiler, J. M., Ansell, J. M., Anstice, N. S., Chase, J. G., Gamble, G. D., . . . Paudel, N. (2015). Neonatal glycemia and neurodevelopmental outcomes at 2 years. *New England Journal of Medicine, 373*(16), 1507-1518. doi:10.1056/NEJMoa1504909

McKinlay, C. J., Alsweiler, J. M., Anstice, N. S., Burakevych, N., Chakraborty, A., Chase, J. G., . . . Jiang, Y. (2017). Association of neonatal glycemia with neurodevelopmental outcomes at 4.5 years. *JAMA pediatrics, 171*(10), 972-983. doi:10.1001/jamapediatrics.2017.1579

Mesholam-Gately, R. I., Giuliano, A. J., Goff, K. P., Faraone, S. V., & Seidman, L. J. (2009). Neurocognition in first-episode schizophrenia: a meta-analytic review. *Neuropsychology, 23*(3), 315. doi:10.1037/a0014708

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., . . . Ross, S. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America, 108*(7), 2693-2698. doi:10.1073/pnas.1010076108

Morgan, A. B., & Lilienfeld, S. O. (2000). A meta-analytic review of the relation between antisocial behavior and neuropsychological measures of executive function. *Clinical psychology review, 20*(1), 113-136. doi:10.1016/s0272-7358(98)00096-8

Posner, M. I., Rothbart, M. K., Sheese, B. E., & Voelker, P. (2012). Control networks and neuromodulators of early development. *Developmental psychology, 48*(3), 827. doi:10.1037/a0025530

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental review, 41*, 71-90. doi:10.1016/j.dr.2016.06.004

Rubia, K. (2013). Functional brain imaging across development. *European child & adolescent psychiatry, 22*(12), 719-731. doi:10.1007/s00787-012-0291-8

Scionti, N., Cavallero, M., Zogmaister, C., & Marzocchi, G. M. (2020). Is cognitive training effective for improving executive functions in preschoolers? A systematic review and meta-analysis. *Frontiers in Psychology, 10*, 2812. doi:0.3389/fpsyg.2019.02812

Sherman, E. M., & Brooks, B. L. (2010). Behavior rating inventory of executive function–preschool version (BRIEF-P): Test review and clinical guidelines for use. *Child neuropsychology, 16*(5), 503-519. doi:10.1080/09297041003679344

Skogan, A. H., Egeland, J., Zeiner, P., Øvergaard, K. R., Oerbeck, B., Reichborn-Kjennerud, T., & Aase, H. (2016). Factor structure of the Behavior Rating Inventory of Executive Functions (BRIEF-P) at age three years. *Child neuropsychology, 22*(4), 472-492. doi:10.1080/09297049.2014.992401

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: a meta-analysis and review. *Psychological bulletin, 139*(1), 81. doi:10.1037/a0028727

Spiegel, J. A., Lonigan, C. J., & Phillips, B. M. (2017). Factor structure and utility of the Behavior Rating Inventory of Executive Function—Preschool Version. *Psychological assessment, 29*(2), 172. doi:10.1037/pas0000324

Zelazo, P. D., Müller, U., Frye, D., Marcovitch, S., Argitis, G., Boseovski, J., . . . Sutherland, A. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development*, i-151. doi:10.1111/j.0037-976x.2003.00260.x