Title	Task utility and norms for the Preschool Executive Task Assessment (PETA)
Authors(s)	Downes, Michelle, Berg, Christine, Kirkham, Fenella J., et al.
Publication date	2017-05-31
Publication information	Downes, Michelle, Christine Berg, Fenella J. Kirkham, and et al. "Task Utility and Norms for the Preschool Executive Task Assessment (PETA)." Taylor & Francis, May 31, 2017. https://doi.org/10.1080/09297049.2017.1333092.
Publisher	Taylor & Francis
Item record/more information	http://hdl.handle.net/10197/10305
Publisher's statement	This is an Accepted Manuscript of an article published by Taylor & Francis in Michelle Downes, Christine Berg, Fenella J Kirkham, Laura Kischkel, Imogen McMurray & Michelle de Haan (2018) Task utility and norms for the Preschool Executive Task Assessment (PETA), Child Neuropsychology, 24:6, 784-798, DOI: 10.1080/09297049.2017.1333092 on 31 May 2017, available online: http://www.tandfonline.com/10.1080/09297049.2017.1333092
Publisher's version (DOI)	10.1080/09297049.2017.1333092

Task utility and norms for the Preschool Executive Task Assessment (PETA)

Authors:

Michelle Downes[1,2], Christine Berg[3], Fenella J Kirkham[2], Laura Kischkel[2], Imogen McMurray[2], Michelle de Haan[2]

Institute(s):

1 School of Psychology, University College Dublin, Dublin, Ireland

2 Developmental Neurosciences, UCL Great Ormond Street Institute of Child Health, London, UK

3 Washington University in St. Louis, MO, USA

Address correspondence to:

Michelle Downes

School of Psychology, University College Dublin, Dublin 4, Ireland

Michelle.downes@ucd.ie

Word Count Abstract: 142

Word Count Main text: 4704

Acknowledgments:

The authors would like to thank the families and teachers who participated in this research. This research was funded by the Child Health Research Charitable

Incorporated Organization and supported by the National Institute for Health Research Biomedical Research Centre at Great Ormond Street Hospital for Children NHS Foundation Trust and University College London. MdH was supported by Great Ormond Street Hospital Children's Charity. A UCL Grand Challenge Award funded the supplementary video. Task materials, including training video and manual, are available by request to first author. The authors have no conflict of interest.

Abstract

Earlier identification of executive deficits in preschool children using an ecological approach would give more scope for intervention. The Preschool Executive Task Assessment (PETA) was developed to resemble an everyday age-appropriate task in order to examine the self-direction and integration of executive functions during a multi-step task. It was designed so that performance can be evaluated in a micro-analytic way and so individualized feedback and support can be easily communicated. The utility of the PETA was assessed with 166 three-to five-year-olds. Results showed improved performance with increasing age and verbal IQ as well as good task reliability and utility. Evidence for influence of socioeconomic status, gender, and use of self-talk was also observed. Clinical applications and future directions of this novel measure are discussed.

Key Words: Executive function, Neuropsychological assessment, Preschool, Ecological assessment, Neurodevelopmental disorders

Abbreviations: PETA=Preschool Executive Task Assessment; EF=Executive Functioning

Introduction

Research suggests that, throughout infancy and the preschool years, performance on executive tasks can be predictive of future executive function (EF) and school readiness (Blair, 2002; Clark, Prior, & Kinsella, 2002; Kraybill & Bell, 2013). EF is susceptible to environmental and disease factors (Hughes & Ensor, 2009; Stuss & Alexander, 2000). However, unlike intellectual quotient (IQ), executive skills show potential for improvement with intervention (Diamond & Lee, 2011) and are an important protective factor in children with low IQ or deficits in other areas (Greenberg, 2006; Johnson, 2012). It is therefore important to identify executive deficits as early as possible. Whether specific executive skills are already established in preschoolers or emerge from a more undifferentiated system with development is still debated (McAuley & White, 2011; Miyake et al., 2000). However, several researchers have found evidence for a unitary construct that becomes more differentiated over time (Senn, Espy, & Kaufmann, 2004; Wiebe et al., 2011). Investigating EF in the preschool years is thus important to understand whether higher-order deficits observed at later ages reflect an early-existing deficit or one that emerges as result of the differentiation within the EF systems that occurs with development.

Measuring executive functioning in the preschool years

The preschool years pose a particularly demanding challenge for researchers due to the rapid development of executive skills from three to five years (Espy et al., 1999). Preschoolers' ability to attend to tasks is more variable than older children, which can make for poor reliability across different measures (Anderson, 2002; Mahone, 2005). Task instructions used for older children can be too linguistically demanding, which is why most behavioural measures of EF, such as the Developmental

Neuropsychological Assessment and the Delis-Kaplan Executive Function System, are standardized for use with children over five years (Delis, Kramer, Kaplan, & Holdnack, 2004; Visu-Petra, Benga, & Miclea, 2007). The one subtest in the NEPSY, which can be administered to children younger than five years, is the Statue subtest that assesses impulse control. Valid assessment of executive skills can be achieved using performance-based assessments in this age range as long as appropriate vigilance is used in the design and development process (Mahone, 2005). There has been a growing interest in preschool executive skills over the past decade, with both new behavioral tasks being developed and tasks from other age ranges being modified to be more age appropriate to account for shorter attention spans, as well as less language and lower motor fluency (Garon et al., 2008). Important design characteristics need to be considered when developing or adapting measures for ease of use with preschool children. These include response modality, language requirements, length of task, ease of use, and novelty (Isquith et al., 2005). The emerging focus on executive assessment in preschool children in the past few years is demonstrated through the development of new batteries such as the Early Childhood Attention Battery (ECAB), NIH Toolbox Cognition Battery and the Preschool Executive Function Battery (Breckenridge, Braddick, & Atkinson, 2013; Garon, Smith & Bryson, 2014; Zelazo et al., 2013) as well as the addition of executive subtests on working memory and processing speed to the latest version of the Wechsler Preschool and Primary Scales (WPPSI-IV, Watkins & Beaujean, 2014).

Preschool assessments that rely solely on summary scores are considered to be questionable as an accurate indicator of cognitive abilities in this age group (Carlson, 2005). Ecologically valid measures that adopt a micro-analytic approach incorporating quantitative and qualitative scoring could better represent specific skills developing at various rates and enhance the diagnostic utility of tasks in this age range (Pritchard & Woodward, 2011). Ecological validity is the extent to which performance on a cognitive task reflects real-life performance and can be measured by a task's verisimilitude and veridicality (Chaytor & Schmitter-Edgecombe, 2003). Verisimilitude concerns the similarity between task demands and everyday demands, while veridicality is the strength of association between task performance and everyday functioning. An ecologically valid task is more engaging for a young child and may reflect real-life performance to a greater extent, providing more accurate information on the type and amount of support a child requires in everyday life. A task that includes a global measure as well as a microanalysis of performance is particularly useful in light of EF models that suggest that subcomponents within this system become more differentiated with age (Hughes, Ensor, Wilson, & Graham, 2010; McAuley & White, 2011).

Ecologically valid measurement of executive functioning

Ecologically valid measures that reflect real-life tasks are important additions to assessment protocols to establish the cognitive and behavioral implications of a child's executive difficulties in daily life (Anderson, 2002; Chan, Shum, Toulopoulou, & Chen, 2008). However, many tasks in the preschool-age range are aimed at measuring specific skills, such as subtests that tap working memory, inhibition, or cognitive flexibility in relative isolation, rather than capturing EF in a more ecologically valid sense where multiple EF components are simultaneously engaged (Garon, Smith, & Bryson, 2014). In regards to intervention, it is important to understand what specific cognitive deficits are present, but to also gain an understanding of the amount and type of support a child requires in a typical activity (Henry & Bettenay, 2010). Tasks in everyday life require the integration of executive skills, unlike the artificial separation of cognitive domains that often occurs in neurocognitive assessments. Rocke and colleagues developed the Children's Kitchen Task Assessment (CKTA) to assess EF in a novel, but ecologically valid, task for eight to 12-year-olds (Rocke, Hays, Edwards, & Berg, 2008). The task is novel in that the child would not have completed this specific task (making play dough from a written recipe) before but is ecologically valid as it is similar in concept to tasks that a child would be expected to do in everyday life (follow instructions to make a toy or complete a school assignment). This cue-based task requires children to follow step-by-step instructions with as little assistance as possible. It evaluates a child's EF and identifies how much support is needed and when this support may be required. Poorer EF was found in school-age children with sickle cell disease on the CKTA (Berg, Edwards, & King, 2012), with significant group differences observed for initiation, organization, and completion. These children also obtained poorer reports on the Behavior Rating Inventory of Executive Function (BRIEF) and the Delis-Kaplan Executive Function System (Delis, Kramer, Kaplan, & Holdnack, 2004; Gioia, Isquith, Guy, & Kenworthy, 2000). In their conclusion, the authors emphasize the importance of triangulation and the richness of information gleaned from using different types of assessment to develop targeted EF support applicable to everyday situations. The aim of this study is to establish the utility of a novel ecological EF measure for use with preschool children.

Development of the Preschool Executive Task Assessment (PETA)

The PETA was developed to measure EF in an ecologically valid way (Burgess et al., 2006; Schmuckler, 2001). Design requirements included a scoring system that is not dependent on linguistic ability or motor speed and accuracy, and a focus on process rather than accuracy so that a young child's executive skills can be captured in a way that mirrors a novel, multi-step task that they may encounter in the classroom. The PETA can be scored both quantitatively and qualitatively for different dimensions of interest and is not time limited. The purpose of a combined cueing system is to highlight a child's strengths and weaknesses as well as creating an objective way to reflect upon where a child's overall performance lies in relation to their peers. The main purpose of the PETA task design is for it to be administered alongside proxy-reports of EF and more traditional lab-based measures in order to gain a more holistic picture of a child's application of EF skills in an everyday context. Information gleaned from task administration is important for translating targeted examples of individual support back to teachers and caregivers. This includes how much and the type or level of cueing an individual child requires in everyday contexts such as a level 1 general verbal cue ("What do you need to get started?") or a level 5 cue of physical demonstration. Importantly, this also includes information on where an individual child struggles in a multi-step task-is it that they need a longer time to complete the task, do they find it difficult to initiate a task independently, do they have difficulty sequencing or moving from one step to another, or is it that they are distractible or poorly organized?

Method

Participants

Testing occurred in three schools across London and at the London Babylab, University College London (UCL) Great Ormond Street Institute of Child Health. In total, 166 children were recruited and assessed (mean age=4.5; range=3.0-6.0; 87 males). Twenty-seven children attended the London Babylab and were recruited through local advertisement. The remaining 139 children were assessed in a quiet setting in their school. Parents of children between 3.0 and 6.0 years of age in the participating schools were sent home study packs containing information sheets and consent forms. Parent consent rates for the 139 children recruited through schools were consistent with previously quoted norms for similar studies at 19/38 (50%) from a Central London school, 46/120 (38%) from an East London school, and 82/260 (32%) from a South London school. Exclusionary criteria included a history of a developmental disorder and not being fluent in English, as determined by parental report.

TABLE 1 HERE

Measures

Demographic information such as age, ethnicity, and gender were collected from parents. Socio-economic status (SES) was based on neighbourhood postcode (school postcode for those assessed on school sites) to estimate total house income on a scale from the UK Office for National Statistic (Nation, Cocksey, Taylor, & Bishop, 2010). Children were divided into low SES and high SES groups based on the mean scale score. Children in the low SES group were from households with an estimated average weekly net income of less than £480 per week.

Preschool Executive Task Assessment

The PETA was adapted from the CKTA, using a similar scoring and cueing system (Rocke et al., 2008). The task was designed so that a wide range of functioning can be accommodated, allowing for a more inclusive approach than a time limited or pass/fail task. The task involves using an "ingredients' box with pre-prepared materials, a recipe book, a timer, and cueing/scoring sheets (Figure 1). The child follows a picture recipe book step-by-step, using the supplied materials, to make the final picture (Figure 2). The examiner delivers a pre- and post-task questionnaire, times task completion, and follows a cueing protocol (see Table S1). The purpose of the verbally administered questionnaires is to introduce each child to the task in the same way and to gain information on how much support that they feel they would require before the task, as well as how much support they perceived themselves to require after task completion. Furthermore, the pre-task questionnaire establishes whether the child typically makes pictures independently without support, and introduces certain materials that the child may not be familiar with-such as the sand timer-in a structured way. Participants are coded on a series of scores after task completion. This includes quantitative scores including the total summary score (TS; the combination of amount and level of cues required to complete the task), as well as completion time, highest of five levels of support required during the task, total number of cues required (TC), and cues specifically required to initiate, sequence, and complete task. Qualitative scores for working memory, distractibility, organization, emotional control and self-talk are recorded by the examiner based on a descriptive guide provided in the task manual (see Table S2).

The goal of the task is to establish how much support a child requires from Step 1 (initiation) through to Step 10 (completion). At the start of each step, the examiner starts at level 1 of the cueing levels with a general verbal cue (Table S1) if the child is struggling and moves through the levels, with two cues in each level and a 10 second delay between each cue, until a child successfully completes the step. The examiner then starts at the first level of cueing for the next step, and so on. Step two, pasting the green circle, is described here as an example: if the child follows the recipe book for this step with no issues, no cues are required. However if the child is struggling (i.e they are sitting and looking around or are playing with materials but with no clear intention to follow instructions), the examiner waits for 10 seconds and then gives a general verbal cue (i.e "What does the recipe book say you need to do?"). If required, this is followed by a second general verbal cue and then, if necessary, the examiner proceeds to two gestural cues (i.e points to page of recipe book/points to correct item in ingredients box), followed by two direct verbal cues (i.e "You have to put that here"/"The recipe book says the green one goes on first"), then to physical assistance (i.e. takes out the page/places circle in place but does not stick) and finally, to level 5 (does step for participant), before starting at level 1 for the next step.

At certain points during the assessment, the child is required to do tasks that tap aspects of executive function, such as inhibition. For example, they must use a sand timer to help them wait for one minute while the ink from the stamper dries. Importantly, the child is not penalized for motor related support from the examiner.

Further task description and demonstration can be observed in the video in the on-line supplementary section.

(FIGURE 1 and 2 HERE)

BRIEF-Preschool (Gioia, Espy, & Isquith, 2003)

The BRIEF-P consists of 63 items. It is comprised of five subscales that create three broader indexes Flexibility (FI), Emergent Metacognition (EMI), and Inhibitory Self-Control (ISCI) and a General Executive Composite (GEC). Higher scores indicate poorer EF (mean =50, SD=10). BRIEF-P data was collected from the teachers of children who were assessed in school settings (n=139). Reasons for non-completion of questionnaires by 51% of the classroom teachers were reported as heavy workloads and lack of time.

Wechsler Preschool and Primary Scale: WPPSI-III-UK (Wechsler, 2002)

The WPPSI-III-UK is a standardized IQ measure used to obtain performance IQ (PIQ) and verbal IQ (VIQ). For three year olds, the full core WPPSI battery was administered but for four and five year olds, two verbal subtests (information and vocabulary) and two performance subtests (block design and matrix reasoning) were used to prorate PIQ and VIQ (normative mean score =100, SD=15). These four subtests were chosen instead of administration of the full core battery for two reasons. Firstly, the core battery for three year olds only contains four subtests and in order to make the administration time comparable across all ages, the same number of subtests was administered to the older children. Secondly, administering four subtests instead of seven reduced the amount of time required for each child to remain outside of the classroom for the purpose of this study.

Procedure

Ethical and R&D approval was obtained from the National Health Service and UCL Great Ormond Street Institute of Child Health (Ref: 13/LO/0962). The testing session took place at the London Babylab, UCL Great Ormond Street Institute of Child Health for 27 children and in a quiet room in the school setting for 139 children. Written parental consent and child assent were obtained. Children were administered the WPPSI followed by the PETA. The total testing session time was typically between 40 and 60 minutes. Teachers were invited to complete the BRIEF-P for participating children in the school settings. All children completed the PETA and the WPPSI-III-UK. Teachers of 67 children returned BRIEF-Preschool forms.

Statistical Analysis

Data analyses were conducted using SPSS for Mac version 21. The interclass correlation coefficient (ICC) was obtained for reliability analyses. Multivariate analysis of variance (MANOVA), chi-square, pearson's correlations, and linear regression were used to look at the association between age, IQ, BRIEF-P and performance. One-way ANOVA was used to investigate group differences based on qualitative ratings, and gender (male/female). Post hoc comparisons were conducted using Tukey HSD. The supplementary section contains tables that show group frequencies and means for PETA task domains (Table S3-S6).

Results

Reliability

To test inter-rater reliability, ten testing sessions were video recorded and coded by three testers blinded to participant details such as age and IQ to obtain the ICC (Hallgren, 2012). The small proportion of the sample assessed for reliability was due to practical reasons as only participants in the lab setting who consented to video were video recorded. Strong inter-rater reliability was observed (ICC =.93). The videos were re-coded by each tester a week later showing evidence for strong intra-rater reliability (ICC= .88 to .98).

PETA Performance, Age, and IQ

Age was strongly related to performance on all quantitative domains of the PETA (TS, TC, Initiation, Sequencing, Meta-cognition, Completion, Time for Completion; r=-.21 to r=-.65, p<.005), except for judgment/safety. Differences in the TS score by age group category (three, four, and five years) were investigated with univariate ANOVA. Post-hoc tests showed that performance significantly increased with age in line with the rapid development of executive skills reported during this period ($F_{2,164}$= 58.39, p <.001; figure 3). A linear regression found that chronological age predicted 40% of the variance in TS ($F_{1,162}$ = 42.9, p <.001, R^2= .398). Younger children required higher levels of examiner support to complete the task with a chi-square showing that three-year-olds tended to require higher levels of cues (e.g. level 4- physical assistance; where the examiner completes part of a step when child does not respond to cues in level 3-direct verbal support) in comparison to four-year-olds with only a small number of five-year-olds requiring physical assistance (X^2 (8)=56.16, *p* <.001). Chi square analysis also showed significant developmental trends for better performance with age in the qualitative domains of working memory (X^2 (4)=16.30, *p* =.003), organisation (X^2 (4)=12.97, *p* =.01), and emotional lability (X^2 (4)=11.77, *p* =.02), with older children receiving higher scores from the examiner for their performance in each of these domains. A non-significant trend for improvements in distractibility was observed (X^2 (4)=6.19, *p* =.15).

A second linear regression was conducted to investigate the potential influence of IQ on PETA TS. After controlling for age, verbal IQ and performance IQ were entered into the model. Verbal IQ significantly contributed to the model, explaining an additional 7% of variance after accounting for age ($F_{1,162}$ = 71.6, p <.001, R^2= .469) but performance IQ did not significantly contribute to the model.

(FIGURE 3 HERE)

PETA Utility

The PETA TS was compared with the BRIEF-P GEC. A significant association was observed between the PETA TS and the BRIEF-P GEC (r=.47, p<.001, n=67) with children requiring a greater level of number of cues on the PETA also receiving poorer reports of EF by their classroom teacher.

Children were scored as ‘poor’, ‘typical’, and ‘very good’ on each of the four qualitative domains (Emotional Lability, Organization, Working Memory, Distractibility) by the experimenter at the end of the testing session. The qualitative PETA domains were investigated using one-way ANOVAs to determine whether the experimenter categorization of 'poor', 'typical', and 'very good' for each of the four domains corresponded to teacher ratings on the approximate domains of the BRIEF-P. Examiner ratings of Organization during the PETA task showed that the poor PETA group obtained the poorest teacher ratings on the BRIEF-P Plan/Organize domain, followed by the typical group, and the very good group ($F_{2,65}$= 3.09, p =.05). PETA Distractibility categorization was compared with scores on the BRIEF-P Inhibit subdomain, however the model was not significant ($F_{2,65}$=.15, p =.86). The model for Working Memory was also not significant ($F_{2,65}$= 1.87, p =.16), although the poor group had lower mean scores than the other two groups. There were no differences between the Emotional Lability groups ($F_{2,65}$= .41, p =.66) on the Emotional Control BRIEF-P domain.

Influence of self-talk, gender, and socioeconomic status on performance

The influence of self-talk (yes/no), gender (make/female), and SES (low/high) were investigated separately for the TS, TC, and Completion Time. Overall, the use of self-talk had no influence on Completion Time ($F_{1,134}$= .282, p =.596) but those who engaged in self-talk obtained a better TS (M=35.26, SD=23.6) than those who did not (M=47.51, SD=40.5; $F_{1,134}$= 4.52, p =.035). A trend for better TC scores in children who engaged in self-talk (M=22.14, SD=11.3) was also observed when compared with children who did not (M=26.7, SD=17.5; $F_{1,134}$= 3.14, p =.079). There were no differences in the rates of self-talk between age-ranges. When investigated further, self-talk had no effect on performance for the four- and five-year-olds but the three-year olds who did not engage in self-talk were found to require more support to complete the task (TC, t(133)=1.77, *p* =.004; TS, t(133)=2.13, *p*=.003) but there was no group difference for Completion Time. A non-significant trend for gender difference showed that girls tended to receive less cues overall (M=23.8, SD=14.8*)* than boys (M=28.4, SD=16.3; t(162)=-1.71, *p* =.06).A lower TS (M=41.2, SD=36.2) also indicated that girls required less support overall than boys *(*M=51.4, SD=39.7; t(162)=-1.90, *p* =.09), but there was no difference for Completion Time. Poorer performance in the "low SES" group (N=47) was observed for TC (M=30.5, SD=18.9) when compared with children in the 'high SES' group (M=24.5, SD=13.9; t(162)=-2.4; p =.04). Poorer overall TS scores were found in the 'low SES' group (M=58, SD=46.5) in comparison to the 'high SES' group (M=42.1, SD=33.7; t(162)=-2.1; p =.05). Longer Completion Time was also observed for children in the 'low SES' group (M=14.9, SD=4.2) as compared to the 'high SES' group (M=13.5, SD=3.2; t(162)-2.0; p =.04).

Discussion

This study presents evidence for both the utility of the PETA as a developmentally sensitive measure and a useful tool for investigating everyday executive development in both clinical and research settings. There is high rater reliability and significant improvements in task performance between three and five years in line with the current understanding of executive development in preschool-age children (Anderson, 2002). Similar to findings in other reports, an influence of verbal IQ on EF performance was also observed (Vriezen & Pigott, 2002). Vriezen and Pigott interpreted the influence of verbal IQ on EF in their patient population of children with acquired brain injury such that children may use verbal mediation to help regulate their behavior, particularly while planning and organizing their behavior to complete a multi-step task. This may be particularly pertinent to the current study due to the early developmental stage of EF in this age range, specifically in the case of the younger children where those who engaged in more self-talk completed the task with less support. One limitation is the lack of relation between the qualitative PETA domains and the related BRIEF-P subdomains. A lack of association between proxy-report measures of EF and performance-based measures is typical for the validation of similar measures in the literature (Dias & Seabra, 2012; Mahone et al., 2002; McAuley, Chen, Goos, Schachar, & Crosbie, 2010; Ponitz et al., 2008). To some degree, these relations may be impacted by higher rates of variability in the younger children (Anderson & Reidy, 2012). Additionally, differences in task constructs and rater reliability for the BRIEF-P may have affected comparisons as well as recent queries around the validity of aspects of the BRIEF-P (Spiegel, Lonigan, & Phillips, 2016). The trend for children falling into appropriate categorical order on the qualitative domains is promising as the qualitative domains add to the informative quality of the assessment and may capture EF aspects that are not commonly observed by parents. The PETA TS showed a greater developmental trend and a consistent relation with EF as reported on the BRIEF-P. However, the real strength of this measure is the inclusion of the qualitative and quantitative components that allow for microanalysis to establish an individual child's strengths and weaknesses. Further validation for the PETA can also be observed in the relationship between self-talk use and better performance in the younger children (Winsler, Carlton, & Barry, 2000), the trend for girls performing slightly better than boys (Ponitz et al., 2008; Wiebe, Espy,

& Charak, 2008) (Ponitz et al., 2008; Wiebe, Espy, & Charak, 2008), and the impact of lower SES on poorer performance (Noble, Norman, & Farah, 2004).

There is a lack of validated EF measures for preschool-age children (Delis, Kramer, Kaplan, & Holdnack, 2004; Korkman, Kirk, & Kemp, 2007). Proxy-report screeners and specific executive tests that are widely used are limited in terms of gauging the impact of deficits on everyday life (Isquith, Crawford, Espy, & Gioia, 2005; Isquith, Gioia, & Espy, 2004). The authors are not aware of another performance-based ecologically valid task available for preschool children, despite the importance of EF for social skills and school readiness in this transition period (Anderson & Reidy, 2012). This task has the potential advantage over existing measures to provide a more holistic and multi-faceted picture of a child's EF (Espy, 1997). Children simultaneously apply a multitude of executive skills in everyday tasks that may not be adequately measured when diluted in a non-natural context. A particular strength of the PETA lies in the ecological nature of its administration. The utility of the PETA scoring system is the ability to capture children who may perform well on brief standalone tasks but may show performance breakdown in the PETA due to the combination of steps, similar to everyday life. Similarly, it could also identify children who may perform poorly on time-limited laboratory measures. Another benefit is that the administrator can easily translate findings from this task into practical recommendations for individualized support. A shortcoming of this task relates to the information that may be lost on specific EF skills when they are assessed in an integrated task. For this reason, the PETA is not intended to replace tasks that quantify basic and specific EF skills but rather to extend the information gleaned from these measures by observing whether differences in these skills are still notable in the context of a novel self-directed multi-step task that is not time-limited. There is a lack of normed and validated EF batteries for the preschool population and this was a limitation when choosing an appropriate EF measure for task validation. However, there have been some very recent advances in the development of EF batteries targeted at preschool-age children since the start of this study and the validity of the PETA TS and the task subcomponents requires further validation with these newly established batteries (Breckenridge, Braddick, & Atkinson, 2013; Garon et al., 2014; Zelazo et al., 2013). The current study suggests that the PETA is a useful measure of EF for preschool-age children and could complement existing questionnaires and more fractionated performance-based measures (Toplak, West, & Stanovich, 2013).

The representativeness of the study population needs to be considered in terms of the generalizability of the reported norms. Although the sample size was comparable to other preschool assessment norms (Breckenridge, Braddick, & Atkinson, 2013) and recruitment rates were within the expected boundaries for research that requires active parental consent, the children who were not consented to participate are more likely to represent children from ethnic minority and lower SES backgrounds (Shaw et al., 2014). This is an important factor to note, as SES was found to influence EF performance in the current study. Further work with patient groups who have known executive deficits is required to establish the clinical utility of the PETA across different patient populations. Studies that compare performance on the PETA with other recently normed and validated batteries of EF for preschool children that segregate specific components of EF are required. Longitudinal validation is also necessary in order to determine whether children with poorer PETA scores will have poorer scores on well-established tasks of EF validated for school-age children, particularly the ecological CKTA task, which also measures EF using a micro-analytic approach (Rocke et al., 2008). Additional future research aims include the development of PETA profiles for clinical and research use as well as the development of an alternate version of the task. An alternate version is important as re-administration of the same task would introduce practice effects and reduce the novelty factor that is integral to capturing EF in this measure.

REFERENCES

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8*(2), 71-82.
Anderson, P. J., & Reidy, N. (2012). Assessing Executive Function in Preschoolers. *Neuropsychology review, 22*(4), 345-360.
Antenor-Dorsey, J. A. V., Hershey, T., Rutlin, J., Shimony, J. S., McKinstry, R. C., Grange, D. K., . . . White, D. A. (2013). White matter integrity and executive abilities in individuals with phenylketonuria. *Molecular genetics and metabolism, 109*(2), 125-131.
Berg, C., Edwards, D. F., & King, A. (2012). Executive function performance on the children's kitchen task assessment with children with sickle cell disease and matched controls. *Child Neuropsychology, 18*(5), 432-448.
Blair, C. (2002). School readiness: Integrating cognition and emotion in a neurobiological conceptualization of children's functioning at school entry. *American Psychologist, 57*(2), 111.
Breckenridge, K., Braddick, O., & Atkinson, J. (2013). The organization of attention in typical development: a new preschool attention test battery. *British Journal of Developmental Psychology*, *31*(3), 271-288.

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of clinical neuropsychology, 23*(2), 201-216.
Charlton, R. A., Barrick, T. R., McIntyre, D. J., Shen, Y., O'Sullivan, M., Howe, F. A., . . . Markus, H. S. (2006). White matter damage on diffusion tensor imaging correlates with age-related cognitive decline. *Neurology, 66*(2), 217-222.
Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology review, 13*(4), 181-197.
Clark, C., Prior, M., & Kinsella, G. (2002). The relationship between executive function abilities, adaptive behaviour, and academic achievement in children with externalising behaviour problems. *Journal of Child Psychology and Psychiatry, 43*(6), 785-796.
Delis, D. C., Kramer, J. H., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis-Kaplan Executive Function System: an update. *Journal of the International Neuropsychological Society, 10*(02), 301-303.
Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4 to 12 years old. *Science, 333*(6045), 959-964.
Dias, N. M., & Seabra, A. G. (2012). Executive demands of the Tower of London task in Brazilian teenagers. *Psychology & Neuroscience, 5*(1), 63.
Espy, K. A. (1997). The Shape School: Assessing executive function in preschool children. *Developmental Neuropsychology, 13*(4), 495-499.
Espy, K. A., Kaufmann, P. M, McDiarmid, M. D, & Glisky, M. L. (1999). Executive functioning in preschool children: Performance on A-not-B and other delayed response format tasks. *Brain and Cognition, 41*(2), 178-199.

Garon, N., Bryson, S. E, & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological bulletin, 134*(1), 31.
Garon, N., Smith, I. M., & Bryson, S. E. (2014). A novel executive function battery for preschoolers: Sensitivity to age differences. *Child Neuropsychology, 20*(6), 713-736.
Gioia, G. A., Espy, K. A., & Isquith, P. K. (2003). *BRIEF-P: Behavior Rating Inventory of Executive Function--preschool Version [kit]*: Psychological Assessment Resources.

Gioia, G. A., Isquith, P. K., Guy, S. C. & Kenworthy, L. (2000) TEST REVIEW Behavior Rating Inventory of Executive Function, *Child Neuropsychology, 6*(3), 235-238.

Greenberg, M. T. (2006). Promoting resilience in children and youth. *Annals of the New York Academy of Sciences, 1094*(1), 139-150.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutor Quant Methods Psychol, 8*(1), 23-34.

Henry, L. A., & Bettenay, C. (2010). The assessment of executive functioning in children. *Child and adolescent mental health, 15*(2), 110-119.

Hughes, C. H., & Ensor, R. A. (2009). How do families help or hinder the emergence of early executive function? *New directions for child and adolescent development, 2009*(123), 35-50.

Hughes, C. H., Ensor, R. A., Wilson, A., & Graham, A. (2010). Tracking executive function across the transition to school: a latent variable approach. *Developmental Neuropsychology, 35*(1), 20-36.

Isquith, P. K., Crawford, J. S., Espy, K. A., & Gioia, G. A. (2005). Assessment of executive function in preschool‐aged children. *Mental Retardation and Developmental Disabilities Research Reviews, 11*(3), 209-215.

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.

Johnson, M. H. (2011). Interactive specialization: a domain-general framework for human functional brain development? *Developmental cognitive neuroscience, 1*(1), 7-21.

Johnson, M. H. (2012). Executive function and developmental disorders: the flip side of the coin. *Trends in cognitive sciences, 16*(9), 454-457.

Korkman, M., Kirk, U., & Kemp, S. (2007). NEPSY-II: Clinical and interpretive manual. *San Antonio, TX: The Psychological Corporation*.

Kraybill, J. H., & Bell, M. A. (2013). Infancy predictors of preschool and post‐kindergarten executive function. *Developmental psychobiology, 55*(5), 530-538.

Levine, B., Robertson, I. H., Clare, L., Carter, G., Hong, J., Wilson, B. A., . . . Stuss, D. T. (2000). Rehabilitation of executive functioning: An experimental–clinical validation of goal management training. *Journal of the International Neuropsychological Society, 6*(03), 299-312.

Mahone, E. M. (2005). Measurement of attention and related functions in the preschool child. *Mental retardation and developmental disabilities research reviews, 11*(3), 216-225.

Mahone, E. M., Cirino, P. T., Cutting, L. E., Cerrone, P. M., Hagelthorn, K. M., Hiemenz, J. R., . . . Denckla, M. B. (2002). Validity of the behavior rating inventory of executive function in children with ADHD and/or Tourette syndrome. *Archives of Clinical Neuropsychology, 17*(7), 643-662.

McAuley, T., Chen, S., Goos, L., Schachar, R., & Crosbie, J. (2010). Is the behavior rating inventory of executive function more strongly associated with measures of impairment or executive function? *Journal of the International Neuropsychological Society, 16*(03), 495-505.

McAuley, T., & White, D. A. (2011). A latent variables examination of processing speed, response inhibition, and working memory during typical development. *Journal of experimental child psychology, 108*(3), 453-468.

Miyake, A., Friedman, N. P, Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive psychology, 41*(1), 49-100.

Nation, K., Cocksey, J., Taylor, J., & Bishop, D.V.M. (2010). A longitudinal investigation of early reading and language skills in children with poor reading comprehension. *Journal of Child Psychology and Psychiatry, 51*(9), 1031-1039.

Noble, K. G., Norman, M. F., & Farah, M. J. (2005). Neurocognitive correlates of socioeconomic status in kindergarten children. *Developmental science, 8*(1), 74-87.

Ponitz, C. E. C., McClelland, M. M., Jewkes, A. M., Connor, C. M., Farris, C. L., & Morrison, F. J. (2008). Touch your toes! Developing a direct measure of behavioral regulation in early childhood. *Early Childhood Research Quarterly, 23*(2), 141-158.

Pritchard, V. E., & Woodward, L. J. (2011). Preschool executive control on the Shape School task: Measurement considerations and utility. *Psychological assessment, 23*(1), 31.

Rocke, K., Hays, P., Edwards, D., & Berg, C. (2008). Development of a performance assessment of executive function: The Children’s Kitchen Task Assessment. *The American Journal of Occupational Therapy, 62*(5), 528-537.

Sarsour, K., Sheridan, M., Jutte, D., Nuru-Jeter, A., Hinshaw, S., & Boyce, W. T. (2011). Family socioeconomic status and child executive functions: the roles of language, home environment, and single parenthood. *Journal of the International Neuropsychological Society, 17*(01), 120-132.

Senn, T. E., Espy, K. A., & Kaufmann, P. M. (2004). Using path analysis to understand executive function organization in preschool children. *Developmental neuropsychology, 26*(1), 445-464.

Shaw, T., Cross, D., Thomas, L. T., & Zubrick, S. R. (2015). Bias in student survey findings from active parental consent procedures. *British Educational Research Journal, 41*(2), 229-243.

Spiegel, J. A., Lonigan, C. J., & Phillips, B. M. (2016). Factor Structure and Utility of the Behavior Rating Inventory of Executive Function-Preschool Version. *Psychological assessment*.

Stuss, D. T., & Alexander, M. P. (2000). Executive functions and the frontal lobes: a conceptual view. *Psychological research, 63*(3-4), 289-298.

Visu-Petra, L., Benga, O., & Miclea, M. (2007). Dimensions of attention and executive functioning in 5-to 12-years-old children: Neuropsychological assessment with the NEPSY battery. *Cognition, Brain, Behavior, 11*(3), 585-608.

Vriezen, E. R., & Pigott, S. E. (2002). The relationship between parental report on the BRIEF and performance-based measures of executive function in children with moderate to severe traumatic brain injury. *Child Neuropsychology, 8*(4), 296-303.

Watkins, M. W., & Beaujean, A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition. *School Psychology Quarterly, 29*(1), 52-63.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of IntelligenceTM Third Edition (WPPSITM-III)*: Sydney, NSW: Pearson.

Welsh, M. C., & Pennington, B. F. (1988). Assessing frontal lobe functioning in children: Views from developmental psychology. *Developmental neuropsychology, 4*(3), 199-230.

Wiebe, S. A., Espy, K. A., & Charak, D. (2008). Using confirmatory factor analysis to understand executive control in preschool children: I. Latent structure. *Developmental Psychology, 44*(2), 575.

Wiebe, S. A., Sheffield, T., Nelson, J. M., Clark, C. A., Chevalier, N., & Espy, K. A. (2011). The structure of executive function in 3-year-olds. *Journal of experimental child psychology, 108*(3), 436-452.

Winsler, A., Carlton, M. P., & Barry, M. J. (2000). Age-related changes in preschool children's systematic use of private speech in a natural setting. *Journal of Child Language, 27*(3), 665-687.

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner - Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH toolbox cognition battery (CB): measuring executive function and attention. *Monographs of the Society for Research in Child Development, 78*(4), 16-33.

Tables

Table 1. Total group demographics as well as separate demographics for each age range

Variable	Total	3 year olds	4 year olds	5 year olds
Participants tested (N)	166	45	60	61
Male (N, %)	87 (53%)	23	32	32
White British (N, %)	99 (60%)	23	43	33
Black British (N, %)	24 (15%)	7	5	12
Other Ethnic Minority (N, %)	41 (25%)	15	10	16
Low SES	49 (29%)	16	12	21

Performance IQ (M, SD)	101.6(15.9)	103.9(15.5)	102.0(16.3)	99.1(14.7)
Verbal IQ (M, SD)	107.7(16.5)	106.2(18.0)	107.3(16.6)	108.5(16.1)
BRIEF GEC	47.6(10.9)	57.3(12.3)	45.5(9.7)	46.9(10.5)

***p <.005, SES=Socio-economic status

Figure Legends

Figure 1. Illustration of systematically laid out "ingredient box" and child following recipe book to complete step.

Figure 2. A guide to what the steps of the PETA look like as the child progresses through the task. For each step, the child observes a photo of someone completing that step with the required materials as well as a sample image of what they should do. For example, 2(a) shows three pages in the first step of the recipe book that a child observes. The child observes (i) a sample white page with grass on the bottom edge and then (ii) someone using the glue stick on the green circle, followed by (iii) the green circle on the pre-drawn grass in the lower right hand corner of the sample page. They then use the supplied materials in the ingredient box to complete this step before moving onto the next step. At the end of the task the participant observes the final sample picture as demonstrated in 2(b) followed by a STOP sign that indicates that the task is finished. The series of steps that the child completes in the PETA by following the visual recipe book are as follows: (1) Initiation (2) Add green circle (3) Add red circle (4) Add yellow circle (5) Add antennae (6) Add red stamp (7) Let stamp dry for one minute using the timer (8) Draw face and cloud (9) Cut and bend the grass (10) Completion. The sequence of steps is demonstrated in 2(c).

Figure 3. Typical Performance for (a) total number of cues, total score (error bars reflect a 95% confidence interval), and (b) a profile demonstrating the highest level of support required at any stage throughout the task based on age category is illustrated in the above image. Higher scores in the number of total cues and the total score indicate poorer performance in 3(a). Three year olds were more likely to require physical assistance (yellow) or examiner completion (red) on a step during the task but the level of support or cueing at these high levels

decreases for four year olds, and again for five year olds as seen in 3(b). All three year olds required at least some gestural guidance (beige) throughout the steps of the task whereas some four and five years olds completed the task with only general verbal cues (green). No child was completely independent in the successful completion of the task.