A Novel Means-End Problem-Solving Assessment Tool for Early Intervention: Evaluation of Validity, Reliability, and Sensitivity

Andrea Baraldi Cunha, PT, PhD; Iryna Babik, PhD; Natalie A. Koziol, PhD; Lin-Ya Hsu, PT, PhD; Jayden Nord, MA; Regina T. Harbourne, PT, PhD; Sarah Westcott-McCoy, PT, PhD; Stacey C. Dusing, PT, PhD; James A. Bovaird, PhD; Michele A. Lobo, PT, PhD

Department of Physical Therapy (Drs Cunha and Lobo), Biomechanics and Movement Science Program, University of Delaware, Newark, Delaware; Department of Psychological Science (Dr Babik), Boise State University, Boise, Idaho; Nebraska Center for Research on Children, Youth, Families and Schools (Drs Koziol and Bovaird and Mr Nord), University of Nebraska-Lincoln, Lincoln, Nebraska; Department of Rehabilitation Medicine (Drs Hsu and Westcott-McCoy), University of Washington, Seattle, Washington; Department of Physical Therapy (Dr Harbourne), Duquesne University, Pittsburgh, Pennsylvania; Department of Physical Therapy (Dr Dusing), Virginia Commonwealth University, Richmond, Virginia.

Purpose: To evaluate validity, reliability, and sensitivity of the novel Means-End Problem-Solving Assessment Tool (MEPSAT).
Methods: Children with typical development and those with motor delay were assessed throughout the first 2 years of life using the MEPSAT. MEPSAT scores were validated against the cognitive and motor subscales of the Bayley Scales of Development. Intra- and interrater reliability, developmental trends, and differences among groups were evaluated.
Results: Changes in MEPSAT scores positively related to changes in Bayley scores across time for both groups of children. Strong intra- and interrater reliability was observed for MEPSAT scoring across all children. The MEPSAT was sensitive to identify change across time and differences in problem-solving among children with varying levels of motor delay.
Conclusions: The MEPSAT is supported by validity and reliability evidence and is a simple tool for screening early problem-solving delays and evaluating change across time in children with a range of developmental abilities. What this adds to the evidence: The novel MEPSAT is supported by validity and reliability evidence. It is sensitive to detect problem-solving differences among young children with varying motor ability and to capture changes in problem-solving across time. It requires minimal equipment and time to administer and score and, thus, is a promising tool for clinicians to screen for early problem-solving delays or to track intervention progress in young children with or at risk for problem-solving delays. (Pediatr Phys Ther 2021;33:2–9)
Key words: cognition, means-end, motor delay, pediatrics, problem-solving

INTRODUCTION

Means-end problem-solving (MEPS) involves the execution of an intentional sequence of actions performed on a "means" object to achieve a goal related to an "end" object.[1-3] Means-end tasks range in complexity from early developmental tasks like pulling a towel to retrieve a distant object through using tools (ie, using a spoon to eat food, a rake to gather toys, or a key to start a car).[2-6] A MEPS task commonly incorporated in standardized developmental assessments is the means-end towel task, which requires pulling of a towel to retrieve an out-of-reach toy (end object) supported on the towel (means object).[2-5]

0898-5669/110/3301-0002
Pediatric Physical Therapy

Correspondence: Michele A. Lobo, PT, PhD, University of Delaware, 210K CHS Bldg, 540 S. College Ave, Newark, DE 19713 (malobo@udel.edu).

Grant Support: This research was supported by the Institute for Education Sciences (R324A150103).

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.pedpt.com).

The authors declare no conflicts of interest.

DOI: 10.1097/PEP.0000000000000761

MEPS in Children With Typical Development Versus Those at Risk for Motor Delays

Success in MEPS typically improves rapidly after 6 months of age,[3] especially when children become skilled at reaching for and manipulating external tools to achieve their goals.[6,7] The performance of goal-directed behavior in the means-end towel task (eg, pulling the towel and looking at the toy) increases with age after 6 months, leading to greater success in this MEPS task.[4] Children with typical development are usually able to succeed in the means-end towel task at the age of 7 to 8 months.[2-5]

In contrast, children at risk for motor delays often show significant delays in their MEPS abilities.[5,7,8] Risks, such as low socioeconomic status (SES), low birth weight, and prematurity, can negatively affect children's MEPS at early ages.[5,7-10] For example, 6- to 12-month-old children with low SES had delayed performance and less success in the means-end towel task compared with children with moderate to high SES.[5] In two MEPS tasks (towel task and turntable task), young children born preterm (PT) performed more non–goal-directed behavior (eg, looking at the towel/turntable, statically touching or feeling the towel/turntable, mouthing the towel/turntable, and reaching for the distant toy) and less goal-directed behavior than their peers born full term. This was associated with a lower success rate and delayed emergence of intentionality for young children born PT relative to those born full-term.[8] Moreover, young children at risk for autism spectrum disorder (ASD) exhibited fewer problem-solving strategies and required greater assistance in MEPS tasks compared to children who are developing typically.[11] Importantly, delays in MEPS identified in the first year of life often persist throughout the second year of life.[7]

A Novel MEPS Assessment Tool

Means-end problem-solving tasks, such as the means-end towel task, might serve as a marker for identifying early developmental delays in populations with environmental or biological risk factors associated with future learning disabilities.[5,8] Previous research highlighted delays in MEPS performance and learning in young children at risk compared to children who are developing typically.[8] Success in the MEPS task, as well as the amount and types of problem-solving behaviors children exhibit, might serve to identify early learning delays.[4,8] However, there are no MEPS screening tools available to researchers and clinicians. Traditionally, MEPS behaviors have been analyzed in research settings using sophisticated, time-consuming, frame-by-frame behavioral coding of videos.[4,8] Clinically, MEPS tasks might be included as discrete items within more extensive developmental assessment tools.[12]

To simplify the analysis process and provide an affordable, feasible solution for evaluating MEPS in research and clinical settings, we used our extensive MEPS coding experience to develop a novel Means-End Problem-Solving Assessment Tool (MEPSAT). The purpose of this study was to assess construct validity and reliability evidence of scores from the novel MEPSAT administered to young children with typical development or with motor delays. Moreover, we aimed to evaluate whether the MEPSAT is sensitive enough to capture change in problem-solving ability with age and to detect differences among children with varying levels of motor delay. We hypothesized that the MEPSAT would demonstrate adequate validity and reliability evidence, as well as sensitivity.

METHODS

Participants

Fifty-two young children participated in this study: 22 children who are developing typically (13 males; Mean = 5.9 months at the first visit, SD = 0.2) enrolled in a longitudinal observational study,[4] and 30 children with motor delay (17 males; Mean = 10.4 months [prematurity-corrected] at the first visit, SD = 2.4) enrolled in a longitudinal randomized controlled trial.[13] This project was approved and monitored by the Internal Review Boards (IRBs) at Virginia Commonwealth University and Duquesne University (IRBs for all other sites).

Eligibility criteria for children who are developing typically were: (1) born full-term (>37 weeks of gestational age); (2) without gross motor delays evidenced by the Bayley Scales of Infant and Toddler Development, 3rd Edition (Bayley-III[12]); (3) absence of genetic or medical diagnoses; and (4) availability to schedule assessment visits by 6 months of age to capture the emergence of means-end learning in the towel pull task.

Eligibility criteria for children with motor delays were: (1) diagnosis of cerebral palsy (CP), increased risk for CP due to prematurity or perinatal brain damage, or motor delay of an unknown origin; (2) gross motor delay evidenced by a scaled score of more than 1 SD below the mean for the gross motor subscale of the Bayley-III[12,13]; (3) absence of a primary diagnosis of a genetic syndrome, ASD, or spinal cord injury, and medical complications, such as severe visual impairment or a progressive medical disorder (criteria for involvement in the larger randomized intervention trial these children were enrolled in); and (4) 7 to 16 months of corrected age, able to sit with support of their arms for 3 seconds,[13] and not demonstrating means-end learning at the first visit. Children in this group were further classified by the severity of their motor delay (n = 10 mild; n = 10 moderate; and n = 10 significant) based on a scale incorporating their Gross Motor Functional Classification System level,[14] distribution of motor deficit, and active movement observed.[13] Additional sample characteristics are in Table 1.

Procedure

All children were tested in their homes 5 times. Children who are developing typically were assessed at 6, 9, 12, 18, and 24 months of age (visits per child: Mean = 4.5, SD = 0.7;

TABLE 1

Demographic and Health-Related Information for Children Developing Typically and Children With Motor Delays

Criteria	Children Developing Typically	Children With Motor Delay
Sex		
Female	45.5%	43.3%
Male	54.5%	56.7%
Race		
African American	18.2%	16.7%
Caucasian	72.7%	66.7%
Asian	9.1%	8.3%
Mixed race	0.0%	8.3%
Ethnicity		
Hispanic/Latino	0.0%	6.7%
Gross household income		
\$0-\$14 999	4.5%	6.7%
\$15 000-\$24 999	0.0%	10.0%
\$25 000-\$34 999	9.1%	10.0%
\$35 000-\$44 999	4.5%	3.3%
\$45 000-\$59 999	9.1%	3.3%
\$60 000-\$79 999	22.7%	10.0%
≥\$80 000	50.0%	56.7%
Gestational age, wk	39.4 ± 1.1	33.9 ± 6.8
Received early intervention services during the study	0.0%	70.4%

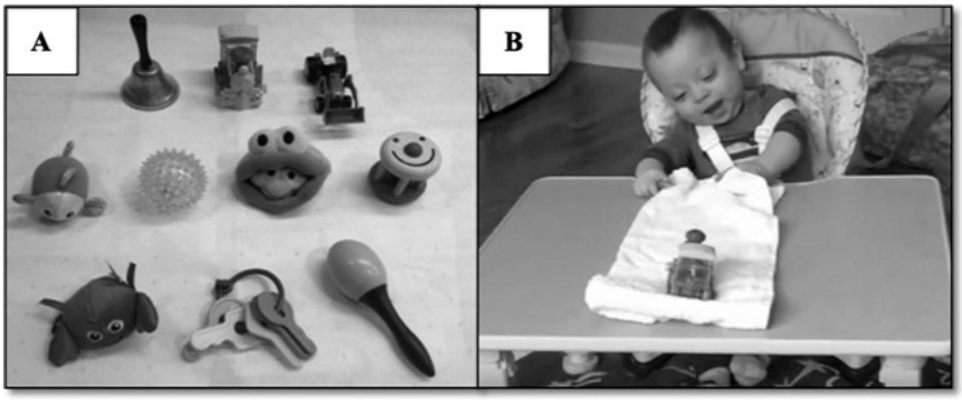

Fig. 1. Example of toys (A) and the screening setup (B) for the Means-End Problem-Solving Assessment Tool.

missing visits: 8.2%). Children with motor delay were assessed at an initial visit (on average, at 10.4 ± 2.4 months) and then 1.5, 3, 6, and 12 months later (visits per child: Mean = 5.0, SD = 0.0; missing visits: 0%). Although the groups were enrolled into studies that differed in the timing of visits, the assessment procedures were identical, the number of visits was the same, and the first visit occurred prior to the emergence of successful MEPS in each group. Thus, these data allowed us to document initial MEPS learning and longitudinal change in groups of children with varied motor ability.[4,8]

At each visit, children engaged in a MEPS task that assessed their ability to pull a towel (means object) to obtain an out-of-reach, supported toy (end object).[2,3] For this task, children were either seated in a portable high chair with tray or on their parent's lap at a surface.[3,4,8] Each child was provided three 30-second trials with a small, interesting toy placed on the far end of the towel, out of the child's reach (Figure 1). The researcher attempted to attract the child's attention to the toy by touching the toy and saying, "Get the toy" at the beginning of each trial. Trials ended when children touched the end object, 30 seconds elapsed, or the child moved the task objects out of reach (eg, dropped them on the floor).[4,8] The visits were video recorded and behaviors were scored from videos using the MEPSAT scoring algorithm (Figure 2; see Supplemental Digital Content 1, available at: http://links.lww.com/PPT/A310).

Children's cognitive and motor development were assessed at each visit (except the 1.5-month visit for children with motor delays) using the cognitive, fine motor, and gross motor subscales of the Bayley-III.[12] Bayley items were scored from videos by a trained assessor blind to children's scores on the MEPSAT. Raw Bayley scores were used for statistical analyses.

Instrument and Outcomes Measures

The MEPSAT was developed based on their experience coding and assessing MEPS in children with typical development and in children with developmental delays. The MEPSAT provides two outcomes: (1) a determination of whether successful MEPS was demonstrated (means-end learning: yes/no; Figure 2A) and (2) a rating of the level of means-end performance (0-9 interval scale; Figure 2B). The determination of whether means-end learning occurred is based on a rubric incorporating the outcome measures: (a) *toy contact*—whether the child contacted the toy by pulling the towel; (b) *looking at the toy*—whether the child looked at the toy in the 5 seconds prior to contacting it; and (c) *intentionality*—whether the child attempted to retrieve the toy, scored as either "no evidence of intention," "unclear/ambiguous intention," or "clear intention." *Means-end learning* occurred when contact was made with the end toy, visual attention to the toy occurred in the 5 seconds before contact, and intention to retrieve the toy was rated as clear for a trial.

Level of means-end performance is rated on the MEPSAT on a scale from 0 through 9, with 9 reflecting the highest level of performance. This is achieved using the decision tree in Figure 2B. The rating is based on the variables described earlier, as well as the occurrence of the following behaviors during each trial: (1) *mouthing the towel*—mouth, tongue, or lips in contact with the towel; (2) *touching the towel*—hand(s) statically touching or feeling the towel, but not moving it; (3) *lifting the towel*—lifting at least half of the towel without pulling it; (4) *reaching for the out-of-reach toy*—extending the arm(s) toward the out-of-reach toy; and (5) *pulling the towel*—pulling or sliding the towel in any direction on the surface.[4,8] The *MEPS levels* are listed and described in Table 2.

Statistical Analyses

For all statistical analyses, data were aggregated across the 3 trials within each visit.

Preliminary Analyses. The equivalence between the 2 groups (children who are developing typically vs those with motor delay) on demographic parameters at the beginning of the study was tested using PASW Statistics software (version 18.0.3) with a χ^2 analysis for sex (0 = males; 1 = females), race (1 = Black; 2 = Caucasian; 3 = Asian; 4 = mixed), gross household income category (values 1-7 according to the income group; Table 1), and the presence of therapy services (0 = no; 1 = yes); and with an independent-samples t-test for gestational age (a continuous variable measured in weeks of gestational age).

Validity Evidence of the Means-End Problem-Solving Assessment Tool Scores. Concurrent construct validity evidence was evaluated by comparing MEPSAT scores (means-end learning and level of means-end performance) with the

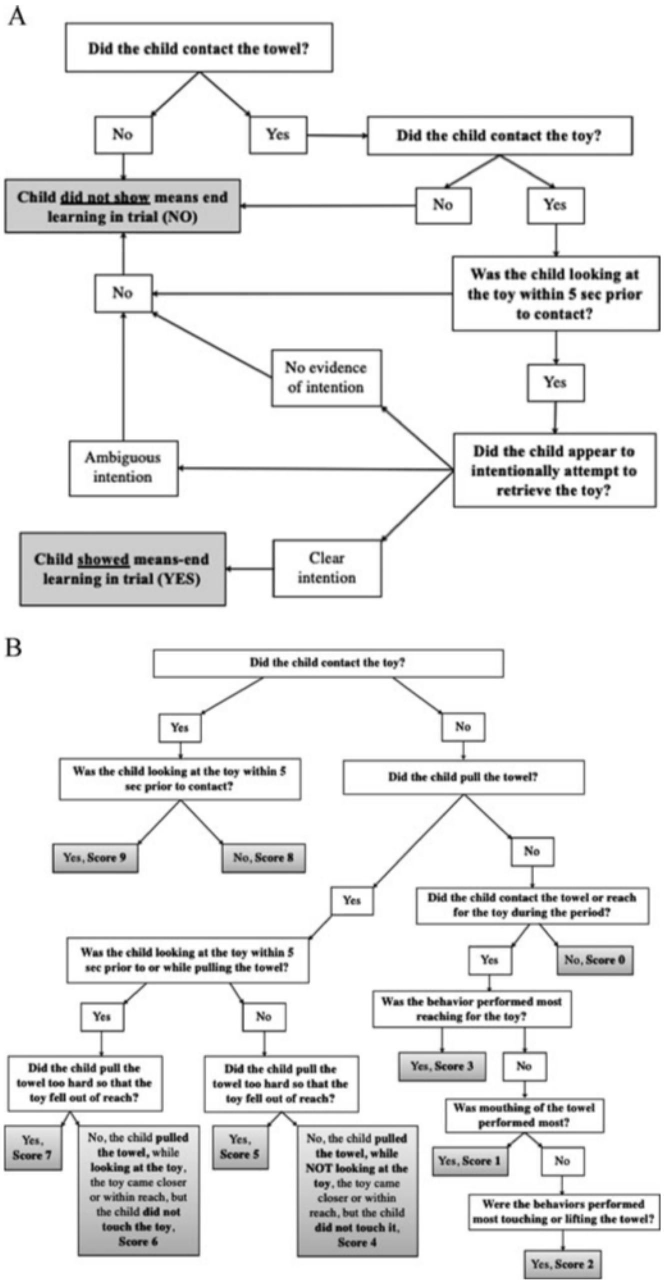

Fig. 2. Means-End Problem-Solving Assessment Tool: (A) determination of means-end learning and (B) rating the level of means-end performance on a scale from 0 through 9.

cognitive, fine motor, and gross motor scores from the Bayley-III in children with typical development and in those with motor delay. We used Bayley *cognitive score* as a criterion because the MEPSAT aims to assess the cognitive construct of means-end understanding. Moreover, Bayley *fine* and *gross motor scores* were used to provide convergent evidence because performance on the MEPS is contingent upon motor ability in the form of postural control, gaze stabilization, reaching, and grasping.[3,4]

Linear mixed modeling (LMM) was performed in SAS (version 9.4)[15,16] to evaluate associations between the MEPSAT and Bayley scores. See Supplemental Digital Content 2 (available at: http://links.lww.com/PPT/A311) for the model parameterization. We estimated the following parameters: (1) *within-child associations*—variation within each child across time, or the extent to which children's variations in MEPSAT scores across time predicted variations in their Bayley scores across time, and (2) *between-child associations*: differences between children while averaging scores across visits, or the extent to which children's MEPSAT scores averaged across time predicted their Bayley scores averaged across time.

Effect sizes were calculated using Cohen's f^2 (0.02 = small, 0.15 = medium, 0.35 = large effect)[17] and standardized regression coefficients.

Reliability Evidence of MEPSAT Scores. Reliability evidence of the MEPSAT was assessed via rescoring 46.2% of the data for intrarater reliability (agreement within a single rater who recoded videos at least 45 days after the initial coding) and 100% of the data for interrater reliability (agreement between 2 independent raters) in children with typical development and those with motor delay.

TABLE 2

Level of Means-End Performance Rated on a Scale From 0 Through 9 on the Means-End Problem-Solving Assessment Tool

Score	Level of Means-End Performance
0	No towel contact
1	Mostly mouthed the towel
2	Mostly touched/lifted the towel
3	Mostly reached for the out-of-reach end toy
4	Pulled the towel without looking at the toy, toy moved closer
5	Pulled the towel too hard without looking at the toy, toy fell out of reach
6	Pulled the towel while looking at the toy, toy came closer
7	Pulled the towel too hard while looking at the toy, toy fell out of reach
8	Contacted the toy without looking at it
9	Contacted the toy while looking at it

For *means-end learning* (4-point ordered categorical variable), quadratic weighted dependent kappas[18] were used to estimate inter- and intrarater reliability coefficients. Dependent kappas were necessary to account for multiple visits nested within children. For *level of means-end performance* (10-point ordered categorical variable), 2-way mixed-effects intraclass correlation coefficients (ICCs)[19,20] were estimated for inter- and intrarater reliabilities. Bootstrapping was used to calculate empirical 95% confidence intervals for the ICCs to account for the multilevel nature of the data. Analyses were performed in R (version 3.6.1) using the multiagree and rel packages.[21-23]

Differences in MEPSAT Scores Between Children With Typical Development and Those With Mild, Moderate, or Significant Motor Delays. General and generalized LMM performed in SAS[24] was used to evaluate developmental trends (change over time) in MEPSAT scores in children who are developing typically and those with motor delays; differences in trends among the 4 groups (typical children, mild delay, moderate delay, and significant delay) were evaluated. See Supplemental Digital Content 2, 2B and 2C (available at: http://links.lww.com/PPT/A311) for the model parameterizations. *Means-end learning* and *means-end performance* were assumed to follow binomial and normal distributions, respectively.

RESULTS

Preliminary Analyses

There was equivalence between children who are developing typically and those with motor delay on sex ($\chi^2(1) = 0.10$, $P = .786$), race ($\chi^2(3) = 2.57$, $P = .464$), and gross household income parameters ($\chi^2(6) = 4.38$, $P = .625$). However, children with motor delay (Mean = 33.91; SD = 6.77) had lower gestational age compared with children who are developing typically (Mean = 39.42; SD = 1.10; $t(31) = 4.39$, $P < .0001$), and had a higher presence of therapy services at baseline ($\chi^2(1) = 26.11$, $P < .0001$).

Validity Evidence of the Means-End Problem-Solving Assessment Tool Scores

Significant ($P < .05$) and positive (ranging from small to large in magnitude) within-child associations were observed between MEPSAT scores (means-end learning and level of means-end performance) and Bayley-III scores for both children who are developing typically and those with motor delay, whereas significant ($P < .05$) and positive (large in magnitude) between-child effects were observed only for children with motor delay (see Table 3 for detailed statistical significance and magnitude of effects). These results indicate that children with motor delay who showed greater MEPSAT learning scores and level of performance on average also had higher Bayley-III scores. For both groups, variations in MEPSAT learning scores and level of performance across time were related to variations in the Bayley-III scores across time.

Reliability Evidence of the Means-End Problem-Solving Assessment Tool Scores

High intra- and interreliabilities of means-end learning and level of means-end performance scores were found for both children who are developing typically and children with motor delay (Table 4).

Differences in the Means-End Problem-Solving Assessment Tool Scores Between Children With Typical Development and Those With Mild, Moderate, or Significant Motor Delay

Change in Means-End Learning Across Time. There was a significant increase in means-end learning scores across time for all children ($\beta = 0.21$, standard error [SE] = 0.04, $P < .001$). The interaction between time and severity was not statistically significant ($F_{(3,164)} = 0.54$, $P = .658$), suggesting that the average change in means-end learning across time did not vary by severity level. However, there were significant severity level differences in intercepts (ie, averaging across time) ($F_{(3,48)} = 6.92$, $P < .001$), where children with significant motor delay scored lower on average than the other 3 groups (typical: $\beta = 3.95$, SE = 0.90, $P < .001$; mild: $\beta = 3.95$, SE = 0.93, $P < .001$; moderate; $\beta = 2.58$, SE = 1.11, $P = .024$). There were no significant differences between children who are developing typically and children with mild or moderate motor delay (mild: $\beta = 0.20$, SE = 0.58, $P = .736$; moderate: $\beta = 1.37$, SE = 0.86, $P = .117$), or between children with mild delay and those with moderate motor delay ($\beta = 1.17$, SE = 0.92, $P = .206$; Figure 3A).

Change in Level of Means-End Performance With Time. There was a significant increase in level of means-end performance across time for all children ($\beta = 0.15$, SE = 0.03, $P < .001$). The interaction between time and severity was not statistically significant ($F_{(3,164)} = 0.48$, $P = .696$), suggesting that the average change in means-end performance across time did not vary by severity level. In contrast, there were significant differences in intercepts across severity levels ($F_{(3,48)} = 12.24$, $P < .001$). There were no significant differences in average scores between children who are developing typically and those with mild delay ($\beta = -0.34$, SE = 0.68, $P = .624$), whereas there were significant differences between all other group combinations, with the less severely delayed group always outperforming the more severely delayed one (typical vs moderate: $\beta = 1.63$, SE = 0.68, $P = .020$; typical vs significant: $\beta = 3.71$, SE = 0.68,

TABLE 3

Associations Between Means-End Problem-Solving Assessment Tool Outcomes (Means-End Learning and Level of Means-End Performance) and Bayley-III

		Within-Infant Effects				Between-Infant Effects			
Sample	**Bayley**	**β**	**SE**	**P**	**f^2**	**β**	**SE**	**P**	**f^2**
Means-end learning vs Bayley scores									
Typical	Cog	0.56	0.09	*<.001*[a]	0.10	0.03	0.04	.559	0.02
Motor delay	Cog	0.17	0.05	*.001*	0.11	0.52	0.13	*<.001*	*0.63*
Typical	FM	0.59	0.09	*<.001*	*0.22*	0.02	0.05	.739	0.03
Motor delay	FM	0.16	0.04	*<.001*	0.14	0.59	0.13	*<.001*	*0.77*
Typical	GM	0.58	0.09	*<.001*	*0.35*	0.00	0.06	.977	0.05
Motor delay	GM	0.22	0.05	*<.001*	*0.20*	0.46	0.14	*.002*	*0.40*
Level of means-end performance vs Bayley scores									
Typical	Cog	0.45	0.10	*<.001*	0.03	0.01	0.04	.831	0.01
Motor delay	Cog	0.10	0.05	.059	0.03	0.60	0.12	*<.001*	*1.15*
Typical	FM	0.47	0.10	*<.001*	0.05	−0.02	0.05	.606	0.00
Motor delay	FM	0.12	0.04	*.008*	0.07	0.68	0.11	*<.001*	*1.49*
Typical	GM	0.49	0.09	*<.001*	*0.18*	0.01	0.07	.926	0.03
Motor delay	GM	0.15	0.05	*.006*	0.08	0.54	0.13	*<.001*	*0.71*

Abbreviations: β, standardized model coefficients; Cog, cognitive scores; f^2, Cohen's effect size coefficients (0.02 = small, 0.15 = medium, 0.35 = large effect); FM, fine motor scores; GM, gross motor scores; SE, standard errors.
[a]Significant scores, as well as medium and large effects, are *italicized*.

$P < .001$; mild vs moderate: $\beta = 1.97$, SE = 0.80, $P = .017$; mild vs significant: $\beta = 4.04$, SE = 0.80, $P < .001$; and moderate vs significant: $\beta = 2.08$, SE = 0.80, $P = .013$; see Figure 3B).

DISCUSSION

This study provided construct validity and reliability evidence of scores from the novel MEPSAT and highlighted differences in MEPSAT scores among children with varied levels of motor ability across time.

Validity Evidence

In terms of construct validity, means-end learning and performance scores from the MEPSAT demonstrated significant and small to large positive within-child associations with the Bayley cognitive, fine motor, and gross motor subscales for both children who are developing typically and those with motor delays. These results represent close alignment between developmental trajectories of the MEPSAT and Bayley scores. Moreover, significant and large positive between-child effects were observed for children with motor delays, indicating that children with motor delays who had higher MEPSAT learning and performance scores also showed greater Bayley-III scores when scores were averaged across the multiple visits. Statistically nonsignificant between-child effects for children who are developing typically might stem from lower variability in their MEPSAT scores, which might be explained by a potential ceiling effect observed for this population.

Close association of the MEPSAT scores not only with cognitive but also with the motor subscales of Bayley-III suggests that the means-end towel pulling task used in the MEPSAT effectively captures both motor (eg, looking at the towel/toy, exploring the towel, and touching the toy) and cognitive (eg, clear intention to pull the towel to retrieve the toy) aspects of children's early development. During early development, motor and cognitive abilities codevelop and highly depend on each other; thus, this early screening tool would, inevitably, capture both aspects of child development.[3,4,25]

The validity evidence for the MEPSAT also suggests that this tool is capable of identifying delays in problem-solving across time in young children with varying levels of ability. Recent research has shown that MEPS delays are evident across a variety of populations at risk for developmental delays. There is evidence of impaired MEPS in young children born preterm, with low SES, or with increased risk for ASD.[5,7,8,11] Because MEPS delays identified in the first year of life have been shown to persist, screening for MEPS differences in the first year of life may effectively identify early and persistent learning impairments in young children at risk for developmental delays.[5,7-10] Thus, the MEPSAT may satisfy the need for an early screening tool for problem-solving and learning disabilities.[8]

TABLE 4

Intra- and Interrater Reliabilities (With 95% Confidence Interval) for Means-End Learning and Level of Means-End Performance Scoring From the Means-End Problem-Solving Assessment Tool

Sample	**Interrater Reliability**	**Intrarater Reliability**
Quadratic weighted dependent kappas for means-end learning		
Typical	0.869 (0.785, 0.952)	0.936 (0.854, >0.999)
Motor delay	0.979 (0.965, 0.993)	0.857 (0.730, 0.985)
Intraclass correlation coefficients for level of means-end performance		
Typical	0.995 (0.992, 0.999)	0.998 (0.995, >0.999)
Motor delay	0.997 (0.995, 0.999)	0.996 (0.992, >0.999)

Reliability Evidence

Reliability evidence of the MEPSAT scores was strong. Multiple independent raters can arrive at similar scores and

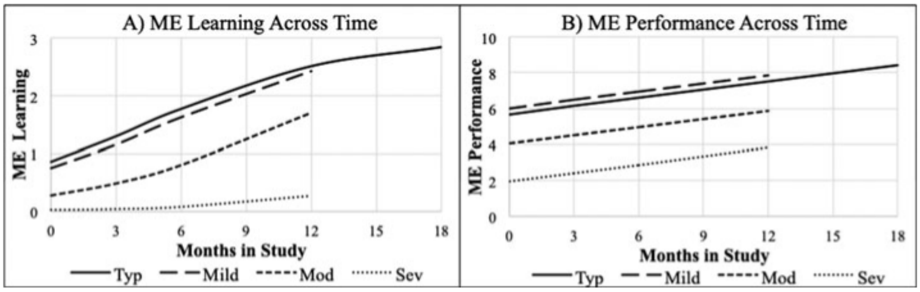

Fig. 3. Means-end learning and performance: estimated developmental trend with time and by severity level; ME = means-end; Typ = typical development; Mild = mild delay; Mod = moderate delay; Sev = severe delay. Note that 0 month in the study occurred on average at 5.9 months of age for the typical group and at 10.4 months' corrected age for the groups with motor delays.

conclusions when using the MEPSAT to determine means-end learning and level of performance for children who are developing typically and those with motor delay. Thus, the MEPSAT might be an appealing screening tool because it provides reliable scores of MEPS in the natural environment, with the potential for use in both research and clinical contexts.[4,7,8] The MEPSAT uses materials commonly available in natural settings (table, towel, and toy) and requires only a few minutes to administer. Trials are brief, so the assessment could be recorded on a smart device for future observation and scoring, if real-time scoring proves challenging. Implementation of this tool may facilitate identification of problem-solving delays and timely provision of intervention to address these delays.

Differences in the Means-End Problem-Solving Assessment Tool Scores Between Children With Typical Development and Those With Mild, Moderate, or Significant Motor Delay

The MEPSAT was sensitive to changes in children's problem-solving across time/age and allowed the charting of developmental trajectories of early MEPS, as young children with typical development and those with motor delay showed positive developmental trends. These results echo previous research suggesting that with age young children perform more goal-directed behavior (eg, looking at the end object and pulling the towel) and increase their problem-solving success.[4,8]

Furthermore, the MEPSAT was sensitive enough to identify differences in means-end learning and performance scores among young children with varying levels of motor ability (typical development or mild, moderate, or significant motor delay). In general, young children with typical development and those with less severe motor delays consistently outperformed more delayed children in both the means-end learning and performance outcomes. Similarly, previous research showed lower success rates and delayed emergence of intentionality for at-risk children compared with children with typical development.[7,8,10,26]

The current findings are in line with previous research showing that postural and trunk control, as well as reaching and grasping abilities, are significantly more impaired for children with increasing severity of motor delay.[27-29] The findings also fit with studies showing that young children at risk for or with CP typically have deficits in trunk control, reaching and grasping skills, and visuomanual coordination,[30-32] all the skills relevant for engaging in general exploration, object manipulation, and MEPS.[4,25] The MEPSAT may be an effective screening tool to provide information about motor ability, cognition, and problem-solving. It may serve as an effective alternative to more labor-intensive methods that have been implemented in prior research.[4,8]

Limitations of the Study

It is important to highlight that even though the developmental trends were similar among children who are developing typically and those with motor delay, timing of the visits differed between the 2 groups. Also, this study tested validity and reliability evidence of the MEPSAT in a modest sample of children varying in motor ability. Future research could investigate the use of the tool in larger samples of children with even more diverse motor and cognitive abilities. Finally, the MEPSAT was scored in this study in real time from video recordings. The next step would be to evaluate the reliability of real-time live scoring.

Clinical Implications

Children with motor delay, especially those with significant motor delays, are at risk for means-end learning and performance delays. The current findings highlight the importance of using adequate tools to screen for early problem-solving delays.[7,8] The MEPSAT is a fast, affordable, and accessible tool backed by initial reliability and validity evidence; thus, it is a promising tool for clinicians to screen for early problem-solving delays or to track intervention progress in young children with or at risk for problem-solving delays.

While the MEPSAT may be used clinically to identify differences in problem-solving among children and across time to demonstrate learning progress in those with motor delay, the MEPSAT does not identify why a child might not be successful in the task. Physical therapists must look at the results of this tool as part of the larger clinical picture for each child. For instance, performance in the task may be limited because of factors that can be mediated through physical therapy intervention, such as limited postural control, poor visual-manual control, or impaired motor planning. Physical therapists can use their evaluation skills to identify these challenges and then address them through intervention. Communication from older, more communicative children can help therapists determine whether the children have trouble understanding how to problem-solve the task versus executing their intended plan. Clinical evaluation might highlight that MEPS impairments are related to factors such as impaired vision that can be addressed by other health professionals.

CONCLUSIONS

The MEPSAT is supported by validity and reliability evidence and is a promising tool for screening early problem-solving delays in young children with a wide range of abilities. It has a simple scoring system that can capture change in children's problem-solving over time, as well as differences among children with varying levels of motor ability. Therefore, the MEPSAT could be used to assess the efficacy of interventions aimed at advancing problem-solving skills, motor abilities, and cognitive outcomes in young children at risk for delays.

ACKNOWLEDGMENTS

The authors would like to thank the participants for their engagement, the research assistants for their efforts coding the data, and Platon Babik for his contributions to the artwork.

REFERENCES

1. Piaget J. *The Origin of Intelligence in the Child*. London, England: Routledge & Kegan Paul; 1953.
2. Willatts P. The stage-iv infant's solution of problems requiring the use of supports. *Infant Behav Dev*. 1984;7(2):125-134. doi:10.1016/S0163-6383(84)80053-3.
3. Willatts P. Development of means-end behavior in young infants: pulling a support to retrieve a distant object. *Dev Psychol*. 1999;35(3): 651-667.
4. Babik I, Cunha AB, Ross SM, Logan SW, Galloway JC, Lobo MA. Means-end problem solving in infancy: development, emergence of intentionality, and transfer of knowledge. *Dev Psychobiol*. 2019;61(2):191-202. doi:10.1002/dev.21798.
5. Clearfield MW, Stanger SB, Jenne HK. Socioeconomic status (SES) affects means-end behavior across the first year. *J Appl Dev Psychol*. 2015;38:22-28. doi:10.1016/j.appdev.2015.02.001.
6. Fagard J, Rat-Fischer L, O'Regan JK. The emergence of use of a rake-like tool: a longitudinal study in human infants. *Front Psychol*. 2014;5:491-491. doi:10.3389/fpsyg.2014.00491.
7. Lobo MA, Galloway JC. Assessment and stability of early learning abilities in preterm and full-term infants across the first 2 years of life. *Res Dev Disabil*. 2013;34(5):1721-1730. doi:10.1016/j.ridd.2013.02.010.
8. Cunha AB, Babik I, Ross SM, et al. Prematurity may negatively impact means-end problem solving across the first 2 years of life. *Res Dev Disabil*. 2018; 81:24-36. doi:10.1016/j.ridd.2018.03.007.
9. Gardner JM, Walker SP, Powell CA, Grantham-McGregor S. A randomized controlled trial of a home-visiting intervention on cognition and behavior in term low birth weight infants. *J Pediatr*. 2003;143(5): 634-639.
10. Petkovic M, Rat-Fischer L, Fagard J. The emergence of tool use in preterm infants. *Front Psychol*. 2016;7:1104-1104. doi:10.3389/fpsyg.2016.01104.
11. Srinivasan SM, Bhat AN. Differences in means-end exploration between infants at risk for autism and typically developing infants in the first 15 months of life. *Dev Psychobiol*. 2019;61(2):203-215. doi:10.1002/dev.21810.
12. Bayley N. *Bayley Scales of Infant and Toddler Development*. 3rd ed. San Antonio, TX: The Psychological Corporation; 2006.
13. Harbourne RT, Dusing SC, Lobo MA, et al. Sitting together and reaching to play (start-play): protocol for a multisite randomized controlled efficacy trial on intervention for infants with neuromotor disorders. *Phys Ther*. 2018;98(6):494-502. doi:10.1093/ptj/pzy033.
14. Palisano R, Rosenbaum P, Walter S, Russell D, Wood E, Galuppi B. Development and reliability of a system to classify gross motor function in children with cerebral palsy. *Dev Med Child Neurol*. 2008;39(4):214-223. doi:10.1111/j.1469-8749.1997.tb07414.x.
15. SAS Institute, Inc. *The SAS system for Windows*. Release 9.4m3. Cary, NC: SAS Institute Inc; 2015.
16. Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger O. *SAS® for Mixed Models*. 2nd ed. Cary, NC: SAS Institute Inc; 2006.
17. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: L. Erlbaum Associates; 1988.
18. Vanbelle S. Comparing dependent kappa coefficients obtained on multilevel data. *Biometrical J*. 2017;59(5):1016-1034. doi:10.1002/bimj.201600093.
19. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods*. 1996;1(1):30-46. doi:10.1037/1082-989X.1.1.30.
20. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428.
21. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019. https://www.R-project.org/
22. Vanbelle S. *Multiagree: Comparison of Dependent Kappa Coefficients*. R package version 3.01, 2019.
23. Martire RL. *rel: Reliability Coefficients*. R package version 1.3.1, 2017. https://CRAN.R-project.org/package=rel
24. Stroup WW. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. Boca Raton, FL: CRC Press; 2013.
25. Lobo MA, Galloway JC. Postural and object-oriented experiences advance early reaching, object exploration, and means-end behavior. *Child Dev*. 2008;79(6):1869-1890. doi:10.1111/j.1467-8624.2008.01231.x.
26. Sun J, Mohay H, O'Callaghan M. A comparison of executive function in very preterm and term infants at 8 months corrected age. *Early Hum Dev*. 2009;85(4):225-230. doi:10.1016/j.earlhumdev.2008.10.005.
27. Hadders-Algra M, van der Fits IBM, Stremmelaar EF, Touwen BCL. Development of postural adjustments during reaching in infants with CP. *Dev Med Child Neurol*. 1999;41(11):766-776.
28. Heyrman L, Desloovere K, Molenaers G, et al. Clinical characteristics of impaired trunk control in children with spastic cerebral palsy. *Res Dev Disabil*. 2013;34(1):327-334. doi:10.1016/j.ridd.2012.08.015.
29. van der Heide JC, Begeer C, Fock JM, et al. Postural control during reaching in preterm children with cerebral palsy. *Dev Med Child Neurol*. 2004;46(4):253-266.
30. Lobo MA, Kokkoni E, Cunha AB, Galloway JC. Infants born preterm demonstrate impaired object exploration behaviors throughout infancy and toddlerhood. *Phys Ther*. 2015;95(1):51-64. doi:10.2522/ptj.20130584.
31. Petkovic M, Chokron S, Fagard J. Visuo-manual coordination in preterm infants without neurological impairments. *Res Dev Disabil*. 2016;51-52:76-88. doi:10.1016/j.ridd.2016.01.010.
32. Saavedra S, Joshi A, Woollacott M, Donkelaar P. Eye hand coordination in children with cerebral palsy. *Exp Brain Res*. 2009;192(2):155-165. doi:10.1007/s00221-008-1549-8.