American Psychological Association

Neuropsychology

Manuscript version of

Identifying Individual Differences in Adolescent Neuropsychological Function Using the NIH Toolbox: An Application of Partially Ordered Classification Modeling

Ryan D. Honomichl, H. Gerry Taylor, Sarah J. A. Carr, Ana E. Istrate, Curtis Tatsuoka

Funded by:

- National Science Foundation

Abstract

Objective: The Cognition Battery of the NIH Toolbox is a commonly utilized set of assessments of neuropsychological abilities, evaluating executive function, attention, working memory, processing speed, and episodic memory. We highlight the utility of an advanced statistical model in providing nuanced characterization of neurocognition in an adolescent population. We propose that partially ordered classification (POSET) models are well suited to analyze polyfactorial tasks and identify distinct profiles of cognitive functioning. ***Method:*** Two models were considered using POSET classification. The first modeled five distinct cognitive functions and allowed for multiple functions to contribute to task performance. The second simpler model involved only two broader-based functions without polyfactorial task specifications. Existing performance data from 745 adolescents aged 14-17 years were analyzed. Posterior probabilities of classification performance and the discriminatory properties of the estimated response distributions indicated how well the modeling approaches fit the data. ***Results:*** The larger first model resulted in eight profiles or "states" characterized by combinations of high or low functioning in five distinct functions. The simpler second model involved two broader-based functions that resulted in four states. Comparing model fit criteria, we believe that the finer-grained first model may better reflect the cognitive constructs associated with the tasks. Notably, POSET modeling did not always provide adequate classification of working memory due to the limited design of the Cognition Battery. ***Conclusions:*** We demonstrate that the use of POSET models is a feasible approach for detailed analysis of neurocognitive data that can extract information on cognitive functions even when provided with limited task batteries.

Public Significance Statement

Our study highlights the utility and importance of advanced statistical modeling methods that systematically account for the multifactorial nature of neuropsychological tasks. Such modeling have the potential to reveal individual profiles of cognitive strengths and weaknesses that account for performance across tests. These profiles can then be examined in association with findings from neuroimaging studies or in guiding individually tailoring educational approaches. Importantly, this analysis illustrates how open-source resources such as the NIH Toolbox can enable application of neuropsychological modeling approaches that require moderate to large samples of test response data, such as POSET modeling.

Keywords: NIH Toolbox, cognitive tests, classification, partially ordered sets, model validation

Identifying Individual Differences in Adolescence Using the NIH Toolbox: An Application of Partially Ordered Classification Modeling

The NIH Toolbox consists of over 100 tasks designed to measure neurological and behavioral function. The goal of the Toolbox is to facilitate the measurement of four domains—Cognition, Sensation, Motor, and Emotion—with common data elements that can be compared across studies (Gershon et al., 2010). The Cognition Battery was constructed to measure a number of areas of function, including executive function, attention, working memory, processing speed, and episodic memory. It can be administered to children as young as 3 years of age using a computerized touchscreen interface (Akshoomoff et al., 2014; Weintraub, Bauer, et al., 2013).

Individual differences in one or more areas can impact developmental outcomes and academic achievement, and so understanding performance in each area is crucial. However, as tests engage multiple neurocognitive functions, it is often difficult to link performance on an individual test with a specific functional outcome. For instance, executive function comprises multiple specific cognitive functions, such as attention and inhibition. However, tasks within the Cognition Battery of the NIH Toolbox are polyfactorial (i.e., typically rely on multiple cognitive functions), making it difficult to infer abilities on any particular function. In our analysis,, we employ novel statistical methods based on partially ordered classification (POSET) models, which are tailored for analysis of polyfactorial tasks to identify detailed profiles of cognitive functioning. Methodological details can be found in Tatsuoka (2002).

Each of the areas evaluated in the Cognition Battery plays an important role in neuropsychological development. Executive function is a reflection of higher order cognitive processes that require top-down control over thought and behavior (Goldstein, Naglieri, Princiotta, & Otero, 2014). Individual differences in this assessment domain have been linked to variation in numerous aspects of psychological and behavioral function (for review, see Diamond, 2013). Neurologically, executive function is most closely associated with prefrontal cortex activity (Miller & Cohen, 2001). In the Toolbox, executive function is assessed in part by the Dimensional Change Card Sort (DCCS) Test, a task that requires the participant to match bivalent test stimuli (e.g., red ball or blue truck) to a target (e.g., red truck or blue ball) based on one dimension (e.g., color) and then, after a number of trials, switch to matching based on the other dimension (e.g., shape). "Switch costs" typically reflect failures in the form of errors or prolonged reaction time due to an difficulty in disengaging attention from the dimensional values of the first rule set. For example, after sorting red balls and blue trucks by color, the individual struggles in inhibiting the focus of their attention on the "blueness" or "redness" of the cards (Kirkham, Cruess, & Diamond, 2003) This phenomenon is similar to task set inertia, found in studies of tasking switching (Keele & Mayr, 2000).

Executive function is also assessed by the Flanker Inhibitory Control and Attention Test. Adapted from the Attention Network Test (Fan, McCandliss, Sommer, Raz, & Posner, 2002; Rueda et al., 2004), the Flanker Test involves responding to the left-right orientation of a central stimulus (e.g., arrow) with additional stimuli on either side of it. Difficulty is manipulated by varying the orientation of these additional stimuli such that they either match (congruent trials) or differ from (incongruent trials) that of the central stimulus. For both the DCCS and Flanker tasks, performance is scored based on both errors made and reaction times. Importantly, both the DCCS and Flanker tasks place demands on attention, another cognitive function of interest here. Broadly speaking, attention refers to the selective or sustained allocation of cognitive focus on or awareness of a stimulus (Langner & Eickhoff, 2013; Posner & Rothbart, 2007).

The Cognition Battery additionally measures working memory, which refers to the ability to monitor and update information in mind (Baddeley, 2012; Smith & Jonides, 1999). Working memory undergoes significant changes during childhood and adolescence (Carriedo, Corral, Montoro, Herrero, & Rucián, 2016; Linares, Bajo, & Pelegrina, 2016) and is associated with academic achievement and developmental outcomes (Gathercole, Pickering, Knight, & Stegmann, 2004; St Clair-Thompson & Gathercole, 2006). In the Cognition Battery, this function is most specifically assessed via the List Sorting Test (Tulsky et al., 2013). During this task, the participant is presented with different pictures and asked to immediately recall the presented list reordered by size. Difficulty level is determined by whether presented objects belong only to one category (e.g., food or animals) or to both.

Processing speed reflects how quickly participants can complete basic cognitive processes. This function improves significantly across childhood and adolescence (Kail, 1991; Kail & Ferrer, 2007) and is also related to educational outcomes, such as component processes of reading (Kail & Hall, 1994) and achievement in mathematics (Geary, 2011). Though processing speed is correlated with other cognitive constructs, such as working memory and IQ (for review, see Bull & Lee, 2014), research confirms associations with academic achievement that are independent of these other skills (Rose, Feldman, & Jankowski, 2011). In the Pattern Comparison Test from the Cognition Battery, processing speed is assessed by having participants view two side-by-side pictures and determine whether they are identical or different. Scores are derived from the number of trials completed within 90 seconds.

Finally, episodic memory reflects the encoding, storage, and retrieval of declarative or explicit information (Squire, 2004). The Picture Sequence Memory Test from the Cognitive Battery assesses episodic memory by presenting the participant with a series of pictured objects and actions related to a thematic event (e.g., working on a farm, going camping), with corresponding audio phrases. Participants must then later select the pictures in the order in which they occur in the event. Task difficulty is manipulated by the number of steps presented (6 to 15)[1]. It is important to note that most if not all of these tests in the Cognition Battery are polyfactorial in that they draw on multiple cognitive functions, and that there is a lack of one-to-one correspondence between task and cognitive function. These polyfactorial relationships are therefore explicitly recognized in the POSET model described below.

Present Study

The NIH Toolbox Cognition Battery has been examined in developmental populations in several previous studies (Akshoomoff et al., 2014; Weintraub, Dikmen, et al., 2013). Our approach applies POSET models (Tatsuoka, 2002; C. Tatsuoka & Ferguson, 2003; Tatsuoka, Varadi, & Jaeger, 2013) to a sample of 14- to 17-year-olds tested on the Cognition Battery as part of an original norming study (Beaumont et al., 2013; Casaletto et al., 2015). POSET classification methods were applied to these data to test a model in which participants are classified into one of a number of discrete profile states based on their performance on the various tasks within the Cognition Battery.

POSET modeling has several advantages. In addition to allowing for the polyfactorial nature of tasks, this approach accesses aggregated response patterns and systematically identifies specific functional domains that may contribute to low or high performance. For example,

[1] In addition to these five functions, the NIH Toolbox Cognition Battery also measures language abilities with a picture vocabulary and reading recognition test. However, these areas are not of interest in the present study, as there is little overlap of the demands of these tasks with those of the other five tasks.

sustained attention, processing speed, and attention control and inhibition all likely contribute to performance on the DCCS and Flanker tasks. If a subject performs well on both, modeling provides evidence that all of these functions, or ability constructs, are at a correspondingly high level (mastery). On the other hand, if a participant performs poorly on both of these tasks, classification would indicate that at least one of the functions is not at a high level. Further analyses examining response patterns across tasks can help identify the source of poor performance. To illustrate, suppose that the participant with poor performance on the DCCS and Flanker tasks performed well on the Pattern Comparison Test, a task that requires both sustained attention and processing speed. In aggregate, these task responses would then support identification of attention control and inhibition as the function being at a low level. The POSET model systematically conducts classification of profiles of cognitive functions in this manner, with modeling extended to more complex scenarios involving multiple tasks and functions.

In brief, POSET models can be considered a systematic statistical analogue to the type of logic used by clinical neuropsychologists. POSET methods explicitly examine the connection between tests and the ability constructs or functions underlying performance of these tests, and have been applied to a variety of patient populations in neuropsychological research, including schizophrenia (Jaeger et al., 2006a,b), Alzheimer's disease (Tatsuoka, et al., 2013), and extreme prematurity (Tatsuoka et al., 2016). POSET models adopt a Bayesian approach in classification. In combining prior probabilities with empirical evidence from test response data, posterior probabilities of "state membership," as characterized by profiles of cognitive functions, are computed to reflect the relative classification evidence for the states being the "true" one for a given subject. Each state has a posterior probability value indicating relative evidence that it is the "true" one, with the sum of the posterior probability values across states adding up to 1.

Classification is to the state with the largest such posterior probability value. The closer this largest value is to 1, the more decisive the classification, and the less the chance the classification is incorrect. Note that one minus the largest posterior probability value is the Bayesian classification error under 0-1 loss function (Berger, 2010), and this is minimized by classifying the subject to the state associated with his or her largest posterior probability value.

Importantly, this approach has an extensive theoretical foundation in terms of the establishment of its statistical properties in classification (Tatsuoka and Ferguson, 2003; Tatsuoka, 2014). A fundamental property of POSET models is their capacity to sift systematically through polyfactorial tests to accurately delineate more nuanced profiles of cognitive functions. This capacity derives from statistical convergence properties that apply when there are sufficient numbers of tests have been administered, in which case the posterior probability value of the "true" state always eventually converges to 1. In other words, with sufficient number of observations (i.e., observed test performances), the "true" state will be classified correctly with high certainty. Indeed, the models have exponential rates of convergence in the sense that the sum of the non-largest posterior probabilities (one minus the largest posterior probability value) converges to zero exponentially. Noting that the NIH Toolbox battery we consider includes only five tests, another appealing property of POSET models is their efficiency in classification of individuals based on performance on brief test batteries.

Another practical advantage of POSET models is that states can be associated with detailed profiles of cognitive functions rather than task performance alone. Profiles thus reflect discrete states for each function assessed by a test battery (e.g., "high" or "low"). These profiles can be partially ordered, in the sense that one state is said to be higher than another if all the first state's subset of cognitive functions subsume those for the second state. Partial ordering of these profiles allows for a first subject to have strengths that a second subject may not, and in turn, for the second subject to have strengths that the first may not. This allows for richer modeling of possible response pattern behavior than unidimensional latent trait models and is theoretically more reflective of neurocognitive diversity within the population of interest.

Adolescents 14 to 17 years of age are of special interest in POSET modeling of the tasks in the Cognitive Battery. Adolescence is a period of rapid change in neurological structure and function (Giedd et al., 1999; Lenroot & Giedd, 2006). Brain growth during this period is accompanied by ongoing development of a variety of cognitive and emotional characteristics, including executive functioning (Steinberg, 2005; Yurgelun-Todd, 2007), attention (Boelema et al., 2014), processing speed (Kail & Ferrer, 2007), and affect processing (Crone, 2009). The development of these characteristics has been linked to important measures of well-being, such as academic achievement (Best, Miller, & Naglieri, 2011; Gathercole, Pickering, Knight, et al., 2004; St Clair-Thompson & Gathercole, 2006; Yu, Branje, Keijsers, Koot, & Meeus, 2013) and social-emotional competence (Liew, 2012; Riggs, Jahromi, Razza, Dillworth-Bart, & Mueller, 2006).

POSET modeling of adolescent performance on the Cognition Battery can offer an appealing alternative to other current modeling approaches of the ability constructs assessed by these tests. Debates exist as to whether the ability constructs assessed by the test battery are independent of one another or are related components of a unified structure (Engle, Kane, & Tuholski, 1999; Kimberg & Farah, 1993; Miyake et al., 2000). A common approach to address this question involves the use of multiple tasks within a sample and the use of factor analysis to extract latent variables. Attempts to examine this issue in adolescents have suggested a multifactorial structure, with the number of factors typically being two or three (Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Shing, Lindenberger, Diamond, Li, & Davidson, 2010). Most commonly, these factors include inhibitory control, working memory maintenance/updating, and set shifting/switching.

POSET modeling is a latent class analytic approach, whereby the individual is categorized or placed into a class or subtype based on profiles of cognitive functions. These functions and their relationship to individual tasks are determined by expert opinion, embodied in a Q-matrix (Tatsuoka, 2009) and then validated analytically in the POSET model framework. As opposed to latent factors linearly determining task response, in POSET models skills are related to response behavior conjunctively, meaning that response or performance distributions on tests depend on whether or not all associated cognitive functions that contribute to test performance are at mastery or not. To perform well on a given test with relatively high probability, a participant is assumed to have relative mastery in all of the functions that contribute to performance. If a participant lacks relative mastery in one of these functions, the response distribution for the test should reflect a lower probability of performing well. The function-specific detail in classification provides additional insights into whether and how the various functions measured by the Cognition Battery are related to one another and which functions are well or poorly measured.

As an illustration, suppose a POSET model has four profiles involving hypothetical cognitive functions A and B. The possible mastery profiles are: both A and B, A only, B only, or neither. The POSET model would appear as a Hasse diagram as in Figure 3. The top state has profile with mastery of functions A and B. The states with mastery of A only and B only are below that top state and are incomparable with each other, in that the subset of mastered functions in one profile are not contained by the corresponding subset in another profile. The state with mastery of neither is the bottom state. Suppose also that tests have binary outcomes, with the probability of “success” being 0.95 given mastery with respect to the associated functions, and the probability of “success” being 0.01 given lack of mastery with respect to at least one of the associated functions. For Bayesian classification, a uniform prior in the POSET model is assumed, so that initially each state is considered as equally likely; in other words, the prior probabilities of state membership are ¼ = 0.25 each. Suppose two tests are administered, a first one that requires function A only and a second one that requires both A and B. If a “success” is observed on the first one, and a “failure” on the second, by using Bayes rule (C. Tatsuoka, 2002), the state with mastery of A only has posterior probability of 0.9517. Observing a “success” on the first test and a “failure” on the second is most likely for subjects in this state, and gives strong indication that it is the true state for that individual. This classification is decisive with only two tasks being administered, which reflects the fast rates of convergence.

In examining the effectiveness of POSET modeling in characterizing cognitive functioning as measured by the NIH Toolbox Cognition Battery, the present study will compare results generated from two models. We applied this novel analytical strategy to characterize the relationship among performances in five cognitive skills, into one of a number discrete states, or hypothetical profiles of function, for each individual in the sample. We also considered an alternative two-function model, where all tests are associated with exactly one function only. The two model fits differ in the granularity of the skills, as well as whether tasks are considered as polyfactorial or unifactorial.

Method

Participants

Participants were adolescents 14-17 years of age drawn from the norming sample for the NIH Toolbox (Beaumont et al., 2013; Casaletto et al., 2016). Of the initial sample of 903 adolescents complete data were available for 745 cases. We rely on complete data cases to facilitate the response distribution parameter estimation process, which depends on Markov Chain Monte Carlo methods (Tatsuoka, 2002). POSET-based classification itself can be applied in incomplete data cases, such as in adaptive testing (Tatsuoka, 2002; Tatsuoka and Ferguson, 2003; Tatsuoka, 2014). Table 1 illustrates the descriptive statistics for both the retained sample and dropped cases. The groups were similar in age, years of maternal education, and sex distribution. However, there were some small discrepancies between retained and dropped cases on race/ethnicity.

Materials and Procedures

Partially ordered set (POSET) modeling and classification methods. Performance from five tests from the NIH Toolbox Cognition Battery were used in the POSET model: the Flanker Inhibitory Control and Attention Test, List Sorting Test, Dimensional Change Card Sort Test, Pattern Comparison Test, and Picture Sequence Memory Test. Details about the construction, makeup, and administration of the tests are available at nihtoolbox.org and in Weintraub et al. (2013). Casaletto and colleagues (2015) examined three types of variables for each test: unnormed, age-normed, and fully-normed data. For fully-normed scores, the authors generated T-scores by creating sub-groups of children in different racial/ethnic groups, with each group’s scores being normalized separately, and with score being regressed on age, maternal education, and sex. After correcting for heterogeneity across groups, these scores reflect one’s neuropsychological performance in comparison to peers matched on age, maternal education, sex, and race/ethnicity. In this study, we base analyses on the fully-normed scores.

POSET modeling requires formulation of a hypothetical model of the neuropsychological functions, or ability constructs, measured by each test. These assumptions, based on clinical judgment by neuropsychologist co-author HGT and congruent with component-theories of executive function and attention (Miyake & Friedman, 2012; Peterson & Posner, 2012; Posner & Rothbart, 2007), are summarized by the "Q-Matrix" presented in Table 1. It is noteworthy that the functions identified in the five-function Q-Matrix overlap somewhat with those areas identified by creators of the Cognition Battery (working memory, processing speed, and episodic memory). However, other functions are distinctly identified (sustained attention, attention control and inhibition). Our characterization of executive function as relating to more specific subprocesses is supported by previous literature that has identified aspects of executive function distinct but correlated with each other (Miyake & Friedman, 2012).

Model validation is assessed by criteria that indicate consistency of data to model specifications, including analysis of 1) the magnitudes of the largest posterior probability values of state membership and 2) the statistical discrimination properties of estimated test response distributions. The first criterion indicates the consistency of observed responses to one profile over all others. Lack of response consistency to any state will generally arise when Q-matrix specifications are incorrect or imprecise. We also computed the sum of the first and second largest posterior probabilities as a reflection of consistency of model to data. The slowest converging non-true states are directly above or below the true state in the partially ordering, and the most difficult to discriminate. These states also have the most similar profiles. So it is of secondary interest to assess these sums as well, especially for short test batteries when there can be a lack of replication for decisive classification.

The second criterion allows for empirical assessment of the extent to which modeled skills map onto test performance as specified in the Q-matrix. Specification of the functions associated with test performance is supported when the resultant partitioning of the states in the POSET model for expected high versus low level performance leads to respective response distribution estimates that are clearly distinguished, and reflect the expected behavior (e.g. subjects at mastery for all associated functions have high probability of performing well, subjects who are not at mastery for all associated functions have high probability of not performing well). Specifically, we compared respective Kullback-Leibler (K-L) information values as the measure of discrepancy between the response distributions within a test, as rates of convergence depend explicitly on these values. Larger K-L information values indicate greater discrepancy between probability distributions, which leads to greater discrimination when applying Bayes rule. The rate of convergence, defined as the asymptotic constant at which the slowest of the non-true state posterior probabilities converges exponentially to zero, are indicative of the difficulty of the classification problem, with slower rates indicating more difficulty. Tatsuoka and Ferguson (2003) and Tatsuoka (2014) show that rates of convergence depend on model complexity around a given state (number of states directly above it in the POSET), as well as the discriminatory properties of the tests, as reflected by Kullback-Leibler information values. Greater model complexity and smaller K-L information values are associated with greater classification difficulty and slower rates of convergence.

In estimation, non-parametric density estimation approaches were adopted, as the shapes of the response distributions appeared complex (Jaeger et al., 2006 & Tatsuoka et al., 2013). Here, we categorized the responses by quartiles, and used conjugate uniform Dirichlet prior probabilities. Gibbs sampling approaches were used with 100,000 iterations after burn-in of

50,000 iterations. Geweke stopping criteria were used in assessing the attainment of stationary convergence among the sampling chains (Geweke, 1992 & Tatsuoka, 2002). Mean and standard deviation of the parameter estimates in the respective multinomial response distributions were computed from the sampled values observed after the burn-in iterations.

Results

Modeling with five cognitive functions in a Q-matrix as in Table 2 yielded a POSET model for the Cognition Battery comprised of eight states. The POSET is depicted as a Hasse diagram in Figure 1. Connections between states indicate ordering, with the higher states indicating more cognitive functions that are intact or at higher levels of mastery compared to the cognitive functions associated with lower states. State 1 (N = 161, 21.61% of the sample) represents the highest state in the model, reflecting high-level mastery of all five cognitive functions, whereas State 8 (N = 107, 14.36%) is the lowest state and indicates a lack of mastery in any function. For the remaining states, at least one function is not at mastery level. State 2 is (N = 124, 16.64%) is characterized by mastery in all skills except working memory and State 4 (N = 85, 11.41%) indicates mastery in all functions except processing speed. In contrast, State 3 (N = 52, 6.98%) represents mastery only in sustained attention, attention control and inhibition, and episodic memory. State 3 thus represents a lower level of cognitive functioning compared with States 2 and 4, as both of the latter states subsume the skills included in the mastery profile for State 3. On the other hand, States 2 and 4 are incomparable, as the functions at mastery level in these profiles do not overlap. State 5 (N = 73, 9.8%) is characterized by three mastered skills (sustained attention, processing speed, and episodic memory), whereas States 6 (N = 28, 3.76%) and 7 (N = 115, 15.44%) represent profiles in which only two functions are at mastery (State 6: sustained attention and processing speed; State 7: sustained attention and episodic memory).

Another noteworthy feature of the model is that for States 5, 6, and 7, the mastery level of working memory is undetermined. In these states, subjects lack mastery of at least one of sustained attention, attention control and inhibition, or episodic memory. Hence, regardless of whether or not a subject has mastery of working memory, the corresponding response distribution for List Sorting is the one for those lacking at least one of those associated functions. As this is the only task in the battery specified as involving working memory in a substantial manner, the model cannot distinguish working memory mastery for subjects in those states. This phenomenon is known as confounding. Explicit recognition of this phenomenon, which is done for POSET models, is essential for the classification model to be identifiable, and for the convergence properties to hold. For State 8, a more accurate description of its profile is that sustained attention is not mastered, and that the rest of the functions are undetermined because all tests are specified as involving sustained attention. If a subject cannot sustain attention, it is difficult to ascertain mastery of any of the other functions. In such instances, we usually assume that all functions are at a low level of mastery.

Posterior Probabilities for State Membership

The mean (standard deviation) of the largest posterior probability for state membership and the sum of the first and second largest values were 0.534 (0.147) and 0.783 (0.127) respectively, indicating that the model specifications were reasonably consistent with patterns of test performance. Figures 2 presents histograms of posterior probabilities for the largest state and the sum of the largest two states. The limited number of tasks in the test battery helps to account for the fact that largest posterior probability values were generally not closer to 1.

Parameter Estimates

Parameter estimates for high and low performers at each quartile are presented in Table 3. High scores reflect good performance for all the tasks. Note that the estimated values support the function-test specifications summarized in Table 2, as there is a clear differentiation between respective response distributions for each test, along with the expected response tendencies. Corresponding Kullback-Leibler information values are listed in Table 3 as well.

Alternative Model

Though the present model provides evidence in support of the five-function model, the quality of the fit of the structure has not yet been compared to an alternative model. In particular, a simpler model may be more appropriate, given the limited number of tasks available. To identify a potential model, we ran an exploratory factor analysis using maximum likelihood with varimax rotation. The results of this analysis suggest a two-function model, with the first function defined by Flanker, DCCS, and Pattern Completion and the second by List Sorting and Picture Sequence Memory. Using this structure in a simple Q-matrix, we identified the first ability construct as “executive function” and the second as “memory function”.

The POSET model for this two-function structure resulted in the identification of four states and depicted as a Hasse diagram in Figure 3. State 1 (N = 263, 35.3% of the sample) is associated with mastery of both executive and memory functions, State 2 (N = 121, 16.24%) with memory function only, State 3 (N = 187, 25.1%) with executive function only, and State 4 (N = 174, 23.36%) with low mastery of both functions.

Figures 4 presents histograms of posterior probabilities for the largest state and the sum of the largest two states. The mean (standard deviation) of the largest posterior probability for state membership and the sum of the first and second largest values were 0.718 (0.169) and 0.940

(0.081) respectively. Parameter estimates for tasks in relation to this model and corresponding K-L information values are presented in Table 4.

Discussion

In the present study, previously published data from a norming study (Beaumont et al., 2013; Casaletto et al., 2016) of the NIH Toolbox Cognition Battery were examined using POSET modeling. Two models were considered, differing in the assumed cognitive specifications of the tests. The first model comprised eight states or profiles of cognitive function, characterized by various combinations of high or low levels of mastery on five functions. The second model consisted of four states and two broader-based functions than were included in the first model and did not assume that any of the tests were polyfactorial. Overall, magnitudes of largest posterior probability values, which indicate classification error in a Bayesian framework, suggest a somewhat better fit for the second, simpler POSET model. However, the eight-state model also fit reasonably well, especially considering the limited number of tests in the battery and added model complexity. Analysis also indicated that the extra states were meaningful in classification, as a sizable number of subjects were classified to every state in the larger model. Statistical discrimination, measured in terms of the Kullback-Leiber information values, was better for the more complex model. Hence, there is support that polyfactorial modeling was nevertheless warranted.

A further finding was that the more fine-grained model did not always provide adequate discrimination of working memory function. The level of mastery of this function could be masked by its lack of differentiation from sustained attention, attention control and inhibition, or episodic memory when one of these functions was at a low level of mastery. These are the other functions associated with the List Sorting test, which is the only test in the Q-matrix associated with working memory. This potential confounding of working memory depends on the profile, as described in Figure 2. This type of confounding can frequently arise with polyfactorial measurement. The smaller model has less model complexity and no such confounding. The reduced complexity of the two-function model appears to be the reason for the higher magnitudes of largest posterior probability values, even with the smaller K-L information values for tests. In other words, the limited size of the Cognition Battery appears to have had a larger impact on classification for the eight-function model than for the two-factor model. Overall, the respective model fits do indicate that the five-function, polyfactorial model is a good fit. With additional measurement, classification at this level of granularity could be improved and the confounding issues with working memory lessened or eliminated.

Previous studies often use a variety of tests with varying demands and task modalities to assess working memory in developmental populations (Alloway, Gathercole, Willis, & Adams, 2004; Gathercole, Pickering, Ambridge, & Wearing, 2004; Huizenga, Crone, & Jansen, 2007). This is not to say that other tasks in the Toolbox fail to place some demands on working memory but only that no other tasks were deemed as having working memory playing an essential role in performance. Granular assessment of functions in POSET models depends in part on the availability of tests in a test battery to discriminate different levels of mastery of a given function. For instance, in Tatsuoka et al (2013), episodic memory was modeled at four levels in a POSET model given its importance in Alzheimer's disease. Future researchers using the Cognition Battery interested in accurately and fully measuring working memory may be well served by supplementing it with other validated measures that do not overlap the same set of functions as specified for the List Sorting test. Adding an attention-focused task also could be helpful in delineating profiles of cognitive function more decisively.

The POSET method has been applied across a range of neuropsychological test batteries (Jaeger et al. 2006a,b; Tatsuoka et al. 2013; Tatsuoka et al. 2016). As noted, when there are polyfactorial tests in the battery, or several gradations of functioning levels within a same cognitive function, it is readily suited to directly reflect such complexities in modeling. Battery design plays a role in successful application, such as having discriminatory tests, sufficient replication, and varied cognitive specifications. The first two characteristics respectively support more decisive classification and the latter one lessens confounding of profiles. Given the time required for cognitive test administration, the number of tests in a battery is often limited due to practical considerations. The NIH Toolbox analysis gives an illustration of the issues that can arise in classification in such instances. Also, as the number of functions being incorporated into the Q-matrix specifications increases, the numbers of possible profiles increases exponentially. Test specifications must be consistent with observed response behavior, which can be harder to satisfy as the models become larger. We have had success with as many as 7 to 8 functions per model. More complex batteries could perhaps employ smaller models in parallel to encompass larger numbers of functions.

A limitation in the scope of the present study is its focus on adolescents aged 14- to 17-year-olds. The norming sample for the NIH Toolbox used in here also contains data from children as young as 3 years and adults as old as 85. The reasons for focusing on adolescents ages 14-17 were twofold. First, the present study is part of a larger project linking cognitive profiles from POSET models to fMRI activation patterns in this age group. Second, adolescence has been shown to be a particularly important time in the development of neuropsychological function (Lenroot & Giedd, 2006; Steinberg, 2005). A focus on this age group is additionally justified by research indicating that cognitive functions become more distinguishable in older youth (Lehto, Juujärvi, Kooistra, & Pulkkinen, 2003; Shing, Lindenberger, Diamond, Li, & Davidson, 2010). Further applications of POSET models are nevertheless warranted to examine developmental changes in the cognitive abilities underlying performance on the NIH Toolbox in younger children.

Conclusion

The open-source nature of the NIH Toolbox Cognitive Battery and its available data make it an important set of measures to consider in conducting neurobehavioral research. We illustrate the utility of the available data in providing nuanced cognitive assessments by developing advanced statistical models of cognitive functioning for a focused age group of 14- to 17-year-olds. POSET models with this number of functions generally require sample sizes of at least 150-200 for there to be sufficient numbers of participants with the various profiles, and to observe a population-representative variety of response patterns. Collecting such data would be difficult and expensive in most cases. Through the approaches described, we obtained a data set of 745 subjects. The subsequent POSET model fit indicates that it is possible, even with a simple battery, to classify individual subject to detailed profiles of cognitive functioning. The model validation approaches focus on specific aspects of model fit, as opposed to “global” tests of goodness of fit. An advantage is that this allows for identification of specific targets for model improvement (Tatsuoka, 2002). This approach can be used to compare Q-matrix specifications more generally. In the future, we plan to associate these detailed profiles with analogous fMRI task activation patterns in a broader neuroimaging study involving adolescents and mathematics learning.

References

Akshoomoff, N., Newman, E., Thompson, W. K., McCabe, C., Bloss, C. S., Chang, L., … Jernigan, T. L. (2014). The NIH Toolbox Cognition Battery: results from a large normative developmental sample (PING). *Neuropsychology*, *28*, 1–10. https://doi.org/10.1037/neu0000001

Alloway, T. P., Gathercole, S. E., Willis, C., & Adams, A.M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, *87*, 85–106. https://doi.org/10.1016/j.jecp.2003.10.002

Baddeley, A. (2012). Working Memory: Theories, Models, and Controversies. *Annual Review of Psychology*, *63*, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Beaumont, J. L., Havlik, R., Cook, K. F., Hays, R. D., Wallner-Allen, K., Korper, S. P., … Gershon, R. (2013). Norming plans for the NIH Toolbox. *Neurology*, *80*(Issue 11, Supplement 3), S87–S92. https://doi.org/10.1212/WNL.0b013e3182872e70

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, *21*, 327–336. https://doi.org/10.1016/j.lindif.2011.01.007

Boelema, S. R., Harakeh, Z., Ormel, J., Hartman, C. a, Vollebergh, W. a M., & van Zandvoort,

M. J. E. (2014). Executive functioning shows differential maturation from early to late adolescence: longitudinal findings from a TRAILS study. *Neuropsychology*, *28*, 177–187. https://doi.org/10.1037/neu0000049

Bull, R., & Lee, K. (2014). Executive Functioning and Mathematics Achievement. *Child Development Perspectives*, *8*, 36–41. https://doi.org/10.1111/cdep.12059

Carriedo, N., Corral, A., Montoro, P. R., Herrero, L., & Rucián, M. (2016). Development of the Updating Executive Function: From 7-Year-Olds to Young Adults. *Developmental Psychology*, *52*, 666–678. https://doi.org/10.1037/dev0000091

Casaletto, K. B., Umlauf, A., Marquine, M., Beaumont, J. L., Mungas, D., Gershon, R., … Heaton, R. K. (2015). Demographically Corrected Normative Standards for the English Version of the NIH Toolbox Cognition Battery. *Journal of the International Neuropsychological Society*, *21*, 378–391. https://doi.org/10.1017/S135561771500137X

Crone, E. a. (2009). Executive functions in adolescence: inferences from brain and behavior. *Developmental Science*, *12*, 825–830. https://doi.org/10.1111/j.1467-7687.2009.00918.x

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual Differences in Working Memory Capacity and What They Tell Us About Controlled Attention, General Fluid Intelligence, and Functions of the Prefrontal Cortex. In A. Miyake & P. Shah (Eds.), *Models of Working Memory* (pp. 102–134). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.007

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, *14*, 340–347. https://doi.org/10.1162/089892902317361886

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The Structure of Working Memory from 4 to 15 Years of Age. *Developmental Psychology*, *40*, 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, *18*, 1–16. https://doi.org/10.1002/acp.934

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology*, *47*, 1539–1552. https://doi.org/10.1037/a0025510

Gershon, R. C., Cella, D., Fox, N. A., Havlik, R. J., Hendrie, H. C., & Wagster, M. V. (2010). Assessment of neurological and behavioural function: the NIH Toolbox. *The Lancet Neurology*, *9*, 138–139. https://doi.org/10.1016/S1474-4422(09)70335-7

Geweke, J. (1992). Evaluating the Accuracy of Sampling-Based Approaches to the Calculations of Posterior Moments. *Bayesian Statistics*, *4*, 641–649.

Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., … Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature Neuroscience*, *2*, 861–863. https://doi.org/10.1038/13158

Goldstein, S., Naglieri, J., Princiotta, D., & Otero, T. (2014). *Introduction: A history of executive* *functioning as a theoretical and clinical construct*. (Goldstein, Sam & J. A. Naglieri, Eds.), *Handbook of Executive Functioning*. Springer. https://doi.org/10.1007/978-1-4614-8106-5

Huizenga, H. M., Crone, E. a, & Jansen, B. J. (2007). Decision-making in healthy children, adolescents and adults explained by the use of increasingly complex proportional reasoning rules. *Developmental Science*, *10*, 814–825. https://doi.org/10.1111/j.1467-7687.2007.00621.x

Jaeger, J., Tatsuoka, C., Berns, S., and Varadi, F. (2006). Distinguishing neurocognitive functions using partially ordered classification models, *Schizophrenia Bulletin*, 32, 679-691.

Jaeger, J., Tatsuoka, C., Berns, S., Varadi, F., Czobor, P., and Uzelac, S. (2006). Associating functional recovery with neurocognitive profiles identified using partially ordered classification models. *Schizophrenia Research*, 85, 40-48.

Kail, R. (1991). Processing Time Declines Exponentially During Childhood and Adolescence. *Developmental Psychology*, *27*, 259–266. https://doi.org/10.1037//0012-1649.27.2.259

Kail, R., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology*, *30*, 949–954. https://doi.org/10.1037/0012-1649.30.6.949

Kail, R. V, & Ferrer, E. (2007). Processing speed in childhood and adolescence: longitudinal models for examining developmental change. *Child Development*, *78*, 1760–1770. https://doi.org/10.1111/j.1467-8624.2007.01088.x

Keele, S. W., & Mayr, U. (2000). Changing internal constraints on action: The role of backward inhibition. *Journal of Experimental Psychology General*, *129*, 4–26. https://doi.org/10.10371/0096-3445.129.1.4

Kimberg, D. Y., & Farah, M. J. (1993). A unified account of cognitive impairments following frontal lobe damage: The role of working memory in complex, organized behavior. *Journal of Experimental Psychology: General*, *122*, 411–428. https://doi.org/10.1037/0096-

Kirkham, N. Z., Cruess, L., & Diamond, A. (2003). Helping children apply their knowledge to their behavior on a dimension- switching task. *Developmental Science*, *6*, 449-467.

Langner, R., & Eickhoff, S. B. (2013). Sustaining attention to simple tasks: a meta-analytic review of the neural mechanisms of vigilant attention. *Psychological Bulletin*, *139*, 870–900. https://doi.org/10.1037/a0030694

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*, 59–80. https://doi.org/10.1348/026151003321164627

Lenroot, R. K., & Giedd, J. N. (2006). Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. *Neuroscience and Biobehavioral Reviews*, *30*, 718–729. https://doi.org/10.1016/j.neubiorev.2006.06.001

Liew, J. (2012). Effortful Control, Executive Functions, and Education: Bringing Self-Regulatory and Social-Emotional Competencies to the Table. *Child Development Perspectives*, *6*, 105–111. https://doi.org/10.1111/j.1750-8606.2011.00196.x

Linares, R., Bajo, M. T., & Pelegrina, S. (2016). Age-related differences in working memory updating components. *Journal of Experimental Child Psychology*, *147*, 39–52. https://doi.org/10.1016/j.jecp.2016.02.009

Miller & Cohen, S., G. E. (2001). An integrative theory of prefrontal cortex function. *Annual*

Reviews Neuroscience, *24*, 167–202.

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions. *Current Directions in Psychological Science*, *21*, 8–14. https://doi.org/10.1177/0963721411429458

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*, 49–100. https://doi.org/10.1006/COGP.1999.0734

Posner, M. I., & Rothbart, M. K. (2007). Research on attention networks as a model for the integration of psychological science. *Annual Review of Psychology*, *58*, 1–23. https://doi.org/10.1146/annurev.psych.58.110405.085516

Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Mueller, U. (2006). Executive function and the promotion of social–emotional competence. *Journal of Applied Developmental Psychology*, *27*, 300–309. https://doi.org/10.1016/j.appdev.2006.04.002

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2011). Modeling a cascade of effects: The role of speed and executive functioning in preterm/full-term differences in academic achievement. *Developmental Science*, *14*, 1161–1175. https://doi.org/10.1111/j.1467-7687.2011.01068.x

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*, 1029–1040.

Shing, Y. L., Lindenberger, U., Diamond, A., Li, S. C., & Davidson, M. C. (2010). Memory maintenance and inhibitory control differentiate from early childhood to adolescence. *Developmental Neuropsychology*, *35*, 679–697. https://doi.org/10.1080/87565641.2010.508546

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science (New York, N.Y.)*, *283*(5408 (Mar. 12, 1999)), 1657–1661. https://doi.org/10.1126/science.283.5408.1657

Squire, L. R. (2004). Memory systems of the brain: A brief history and current perspective. *Neurobiology of Learning and Memory*, *82*, 171–177. https://doi.org/10.1016/j.nlm.2004.06.005

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology (2006)*, *59*, 745–759. https://doi.org/10.1080/17470210500162854

Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, *9*, 69–74. https://doi.org/10.1016/j.tics.2004.12.005

Tatsuoka, C. (2002). Data analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *51*(Part 3), 337–350.

Tatsuoka, C., & Ferguson, T. (2003). Sequential classification on partially ordered sets. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *65*, 143–157. https://doi.org/10.1111/1467-9868.00377

Tatsuoka, C., McGowan, B., Yamada, T., Espy, K. A., Minich, N., & Taylor, H. G. (2016).

Effects of extreme prematurity on numerical skills and executive function in kindergarten children: An application of partially ordered classification modeling. *Learning and Individual Differences*, *49,* 332–340. https://doi.org/10.1016/j.lindif.2016.05.002

Tatsuoka, C., Tseng, H., Jaeger, J., Varadi, F., Smith, M. A., Yamada, T., … Lerner, A. J. (2013). Modeling the heterogeneity in risk of progression to Alzheimer's disease across cognitive profiles in mild cognitive impairment. *Alzheimer's Research & Therapy*, *5*, 14. https://doi.org/10.1186/alzrt168

Tatsuoka, C., Varadi, F., & Jaeger, J. (2013). Latent partially ordered classification models and normal mixtures. *Journal of Educational and Behavioral Statistics*, *38,* 267–294. https://doi.org/10.3102/1076998612458318

Tatsuoka, K. (2009). *Cognitive assessment: an introduction to the rule space method.* New York: Routledge.

Tulsky, D. S., Carlozzi, N. E., Chevalier, N., Espy, K. A., Beaumont, J. L., & Mungas, D. (2013). NIH toolbox cognition battery (CB): Measuring working memory. *Monographs of the Society for Research in Child Development*, *78*, 70–87. https://doi.org/10.1111/mono.12035

Unsworth, N. (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. *Intelligence*, *49,* 110–128. https://doi.org/10.1016/j.intell.2015.01.005

Weintraub, S., Bauer, P. J., Zelazo, P. D., Wallner-Allen, K., Dikmen, S. S., Heaton, R. K., … Gershon, R. C. (2013). NIH toolbox cognition battery (CB): Introduction and pediatric data. *Monographs of the Society for Research in Child Development*, *78*, 1–15. https://doi.org/10.1111/mono.12031

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., … Gershon, R. C. (2013). Cognition assessment using the NIH Toolbox. *Neurology*, *80*(Issue 11, Supplement 3), S54–S64. https://doi.org/10.1212/WNL.0b013e3182872ded

Yu, R., Branje, S., Keijsers, L., Koot, H. M., & Meeus, W. (2013). Pals, problems, and personality: the moderating role of personality in the longitudinal association between adolescents' and best friends' delinquency. *Journal of Personality*, *81*, 499–509. https://doi.org/10.1111/jopy.12027

Yurgelun-Todd, D. (2007). Emotional and cognitive changes during adolescence. *Current Opinion in Neurobiology*, *17*, 251–257. https://doi.org/10.1016/j.conb.2007.03.009

Table 1

Demographic Statistics for Retained Sample and Dropped Cases

	Sample (*N* = 745)	Dropped (*N* = 158)
Age (in years)	15.48 (1.12)	15.46 (1.17)
Mother's education (in years)	12.46 (2.45)	12.26 (2.69)
Gender (% male)	51%	49%
Race/Ethnicity		
White	78.13%	76.35%
Black or African American	20.50%	22.30%
Hispanic	18.93%	17.72%
Asian	2.20%	0.67%
American Indian or Alaska Native	2.34%	2.70%
Native Hawaiian or Pacific Islander	0.14%	0%

Table 2

Q Matrix for Cognitive Tests in NIH Toolbox

	Sustained Attention	Working Memory	Attention Control & Inhibition	Processing Speed	Episodic Memory
Flanker Inhibitory Control and Attention	X		X	X	
List Sorting Test	X	X	X		X
Dimensional Change Card Sort Test	X		X	X	
Pattern Comparison Processing Speed	X			X	
Picture Sequence Memory Test	X				X

Table 3

Parameter Estimates of Task Performance for High and Low Performers for the Five Function Model

	Q1	Q2	Q3	Q4	Kullback–Leibler Information
Flanker Inhibitory Control and Attention Test					
High Performers	0.02 (0.01)	0.2 (0.03)	0.28 (0.03)	0.49 (0.03)	1.19
Low Performers	0.42 (0.03)	0.29 (0.02)	0.22 (0.02)	0.07 (0.02)	
List Sorting Test					
High Performers	0.04 (0.03)	0.15 (0.06)	0.41 (0.07)	0.4 (0.07)	0.81
Low Performers	0.37 (0.02)	0.35 (0.03)	0.17 (0.03)	0.11 (0.03)	
Dimensional Change Card Sort Test					
High Performers	0.02 (0.01)	0.12 (0.03)	0.3 (0.03)	0.56 (0.04)	1.52
Low Performers	0.42 (0.03)	0.34 (0.03)	0.22 (0.03)	0.02 (0.01)	
Pattern Comparison Test					
High Performers	0.12 (0.02)	0.21 (0.03)	0.3 (0.03)	0.37 (0.02)	0.84
Low Performers	0.59 (0.04)	0.28 (0.04)	0.11 (0.04)	0.02 (0.01)	
Picture Sequence Memory Test					
High Performers	0.06 (0.03)	0.23 (0.04)	0.36 (0.04)	0.35 (0.03)	1.13
Low Performers	0.57 (0.06)	0.28 (0.07)	0.09 (0.05)	0.06 (0.04)	

Note: Higher scores indicate better performance

Table 4

Parameter Estimates of Task Performance for High and Low Performers for the Two Function Model

	Q1	Q2	Q3	Q4	Kullback–Leibler Information
Flanker Inhibitory Control and Attention Test					
High Performers	0.03 (0.01)	0.23 (0.03)	0.29 (0.03)	0.45 (0.03)	1.16
Low Performers	0.47 (0.03)	0.27 (0.03)	0.21 (0.02)	0.05 (0.02)	
List Sorting Test					
High Performers	0.09 (0.04)	0.25 (0.04)	0.34 (0.04)	0.32 (0.03)	0.66
Low Performers	0.47 (0.04)	0.33 (0.04)	0.13 (0.03)	0.06 (0.03)	
Dimensional Change Card Sort Test					
High Performers	0.04 (0.02)	0.16 (0.02)	0.32 (0.03)	0.48 (0.03)	1.22
Low Performers	0.46 (0.03)	0.34 (0.03)	0.18 (0.03)	0.01 (0.01)	
Pattern Comparison Test					
High Performers	0.1 (0.02)	0.2 (0.02)	0.27 (0.03)	0.43 (0.03)	0.69
Low Performers	0.49 (0.03)	0.28 (0.03)	0.18 (0.02)	0.05 (0.02)	
Picture Sequence Memory Test					
High Performers	0.06 (0.03)	0.24 (0.04)	0.33 (0.04)	0.36 (0.04)	0.66
Low Performers	0.44 (0.04)	0.26 (0.04)	0.18 (0.03)	0.12 (0.03)	

Note: Higher scores indicate better performance

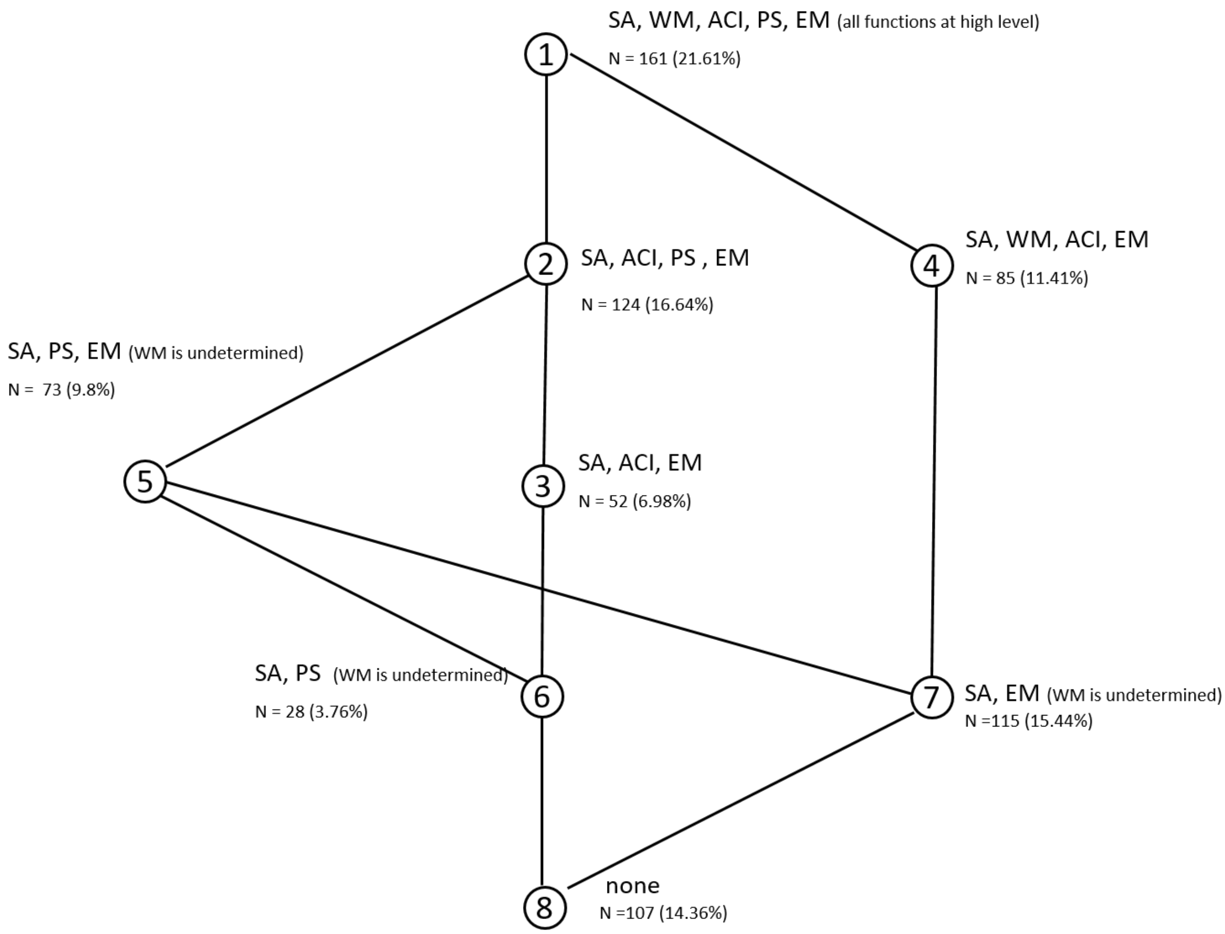

Figure 1. Hasse diagrams of the classification states included in the five function partially ordered set (POSET) model for the NIH Toolbox. High-level functions are listed next to each state number. Abbreviations for each function: SA = sustained attention, WM = working memory, ACI = attention control and inhibition, PS = processing speed, EM = episodic memory.

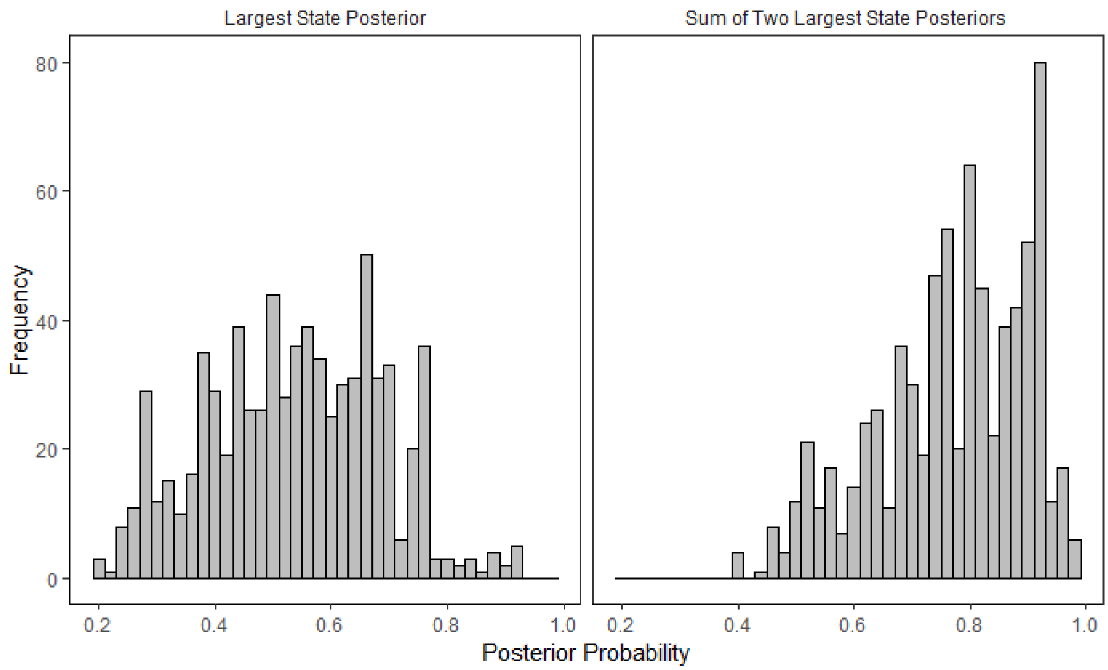

Figure 2. Histograms of the largest and sum of the largest two state posteriors from the five function model.

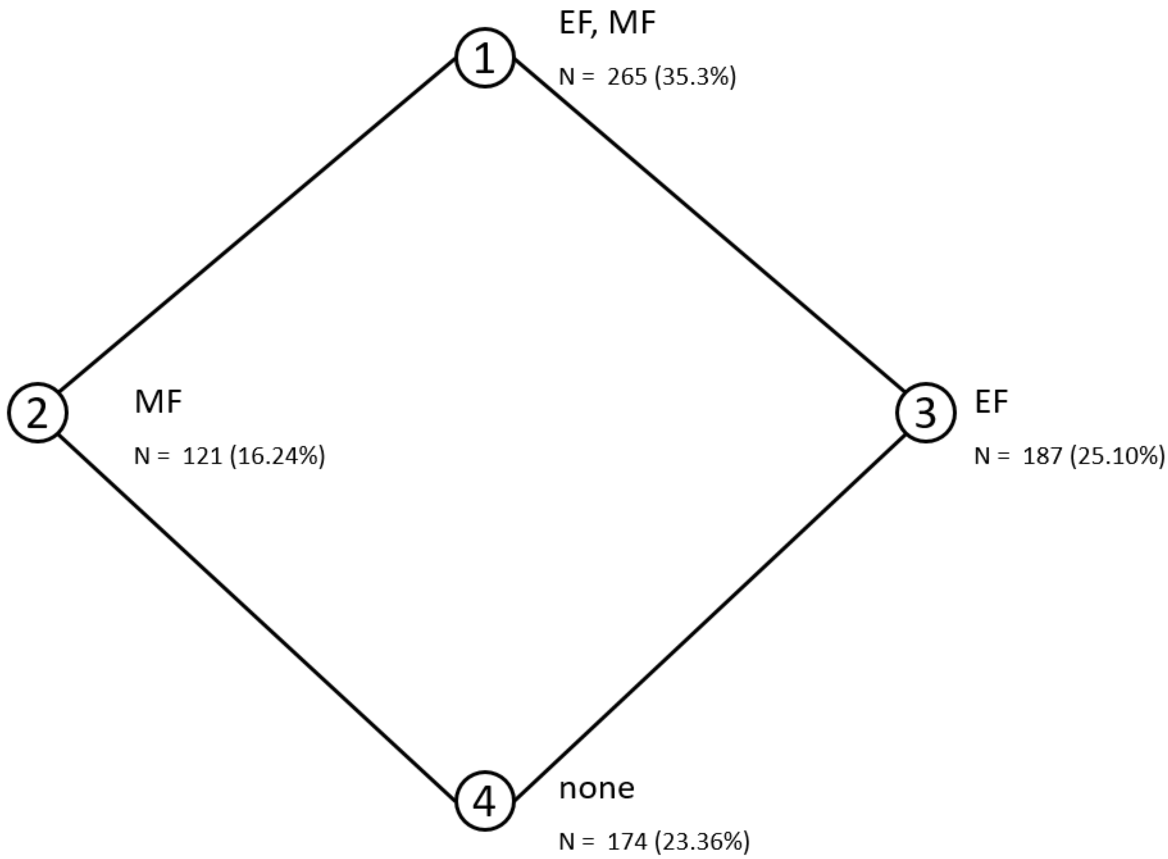

Figure 3. Hasse diagram of the classification states included in the two function POSET model. High-level functions are listed next to each state number. Abbreviations for each function: EF = executive function, MF= memory function.

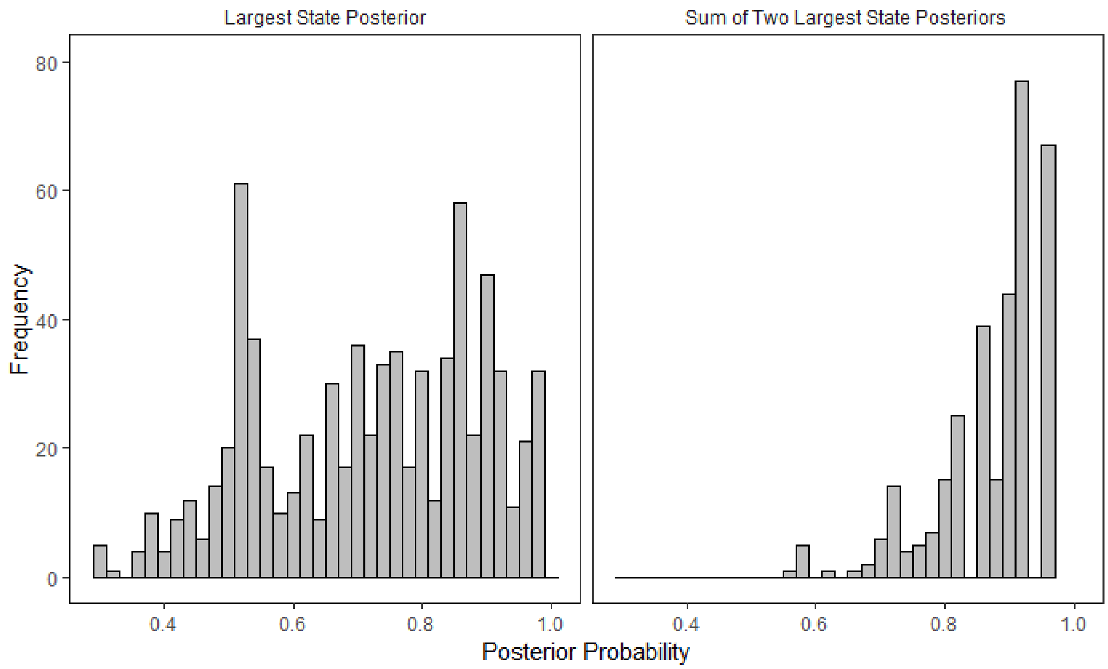

Figure 4. Histograms of the largest and sum of the largest two state posteriors from the two function model.