BRIEF REPORT

CrossMark

Measuring Expectations of Cognitive Enhancement: Item Response Analysis of the Expectation Assessment Scale

Sheida Rabipour[1] · Patrick S. R. Davidson[1,2,3] · Elizabeth Kristjansson[1]

Abstract

Interventions aiming to enhance cognitive functions (e.g., computerized cognitive training and non-invasive brain stimulation) are increasingly widespread for the treatment and prevention of cognitive decline. Drawing on the allure of neuroplasticity, such programs comprise a multi-billion dollar industry catering to researchers, clinicians, and individual consumers. Nevertheless, cognitive enhancement interventions remain highly controversial due to uncertainty regarding their mechanisms of action. A major limitation in cognitive enhancement research and practice is the failure to account for expectations of outcomes, which can influence the degree to which participants improve over an intervention (i.e., the placebo effect). Here, we sought to evaluate the psychometric properties of the Expectation Assessment Scale (EAS), a questionnaire we created to measure the perceived effectiveness of cognitive enhancement interventions. We delivered a web-based version of the EAS probing expectations of either computerized cognitive training or non-invasive brain stimulation. We assessed uni-dimensionality of the EAS using principal component analysis and assessed item properties with a graded item response model. Responses on the EAS suggest good validity based on internal structure, across all subscales and for both computerized cognitive training and non-invasive brain stimulation. The EAS can serve as a reliable, valid, and easily incorporated tool to assess the validity of cognitive enhancement interventions, while accounting for expectations of intervention outcomes. Assessing expectations before, during, and after cognitive enhancement interventions will likely prove useful in future studies.

Keywords Brain training · Cognitive enhancement · Demand characteristics · Expectation · Intervention design · Item response theory · Non-invasive brain stimulation · Placebo effect · Research methods

Introduction

Interventions aiming to enhance cognitive functions are ubiquitous in mainstream society. In particular, techniques involving computerized cognitive training (CCT) and non-invasive brain stimulation (NIBS) are increasingly widespread for the treatment and prevention of cognitive decline. Drawing on the allure of neuroplasticity, such programs comprise a multi-billion dollar industry catering to researchers, healthcare professionals, educators, and individual consumers (SharpBrains 2013, 2016). Nevertheless, cognitive enhancement interventions remain highly controversial due to inconsistent reports of effectiveness and uncertainty regarding their mechanisms of action (Boot and Kramer 2014; Simons et al. 2016).

A major limitation in cognitive enhancement research and practice is the failure to account for expectations of outcomes, which can influence the degree to which participants improve over an intervention (i.e., the placebo effect). The few studies that have examined expectations in the context of cognitive training suggest they may represent an important factor (Boot et al. 2013; Foroughi et al. 2016). No standardized tool exists to measure expectations in the context of cognitive enhancement interventions.

Here, we sought to evaluate the psychometric properties of the Expectation Assessment Scale (EAS), a questionnaire we

✉ Sheida Rabipour
srabi091@uottawa.ca

1 School of Psychology, University of Ottawa, Ottawa, ON, Canada

2 Heart and Stroke Foundation of Ontario Canadian Partnership for Stroke Recovery, Ottawa, ON, Canada

3 Bruyère Research Institute, Bruyère Continuing Care, Ottawa, ON, Canada

created and are using in our intervention research to measure the perceived effectiveness of CCT and NIBS. We report our analyses of participant responses elsewhere (Rabipour et al. 2017).

Materials and Methods

We developed the EAS in two steps: (i) we created an initial version of the EAS based on questions used by Boot et al. (2013); (ii) following an examination of responses on this initial version, we refined the questions to create the current scale (Rabipour and Davidson 2015). As previously described (Rabipour and Davidson 2015), the scale comprises seven items, representing expected outcomes for seven cognitive domains: (i) "general cognitive function," (ii) "memory," (iii)"concentration," (iv) "distractibility," (v) "reasoning ability," (vi) "multitasking ability," and (vii) "performance in everyday activities" (see Supplementary Material). All participants responded based on the same wording, with no additional information provided. Responses were recorded on a 7-point Likert scale.

Participants

We delivered a web-based version of the EAS probing expectations of either CCT (study 1) or NIBS (study 2). Participants were recruited from the student participant pool at the University of Ottawa (UO) and Florida State University (FSU), through ads and flyers (community recruitment), and via web-based recruitment, including Amazon Mechanical Turk (MTurk) and Qualtrics (Table 1). As previously reported (Rabipour et al. 2017), although the majority of participants were recruited solely to complete the survey, the 50 OA recruited from FSU participated in the context of a brain training intervention (Souders et al. 2017).

Data Analysis

Item response theory (IRT) is a "model-based approach" to item and trait analysis (Embretson and Reise 2000). Compared to reliability statistics provided by classical test theory, IRT has numerous advantages, including better estimates of item- and test-level reliability through the marginal reliability for response pattern score. Importantly, it provides standard errors of measurement for each level of the examined trait (for discussion and examples, see Foster et al. 2017; Hays et al. 2000; Kim and Feldt 2010; Shu and Schwarz 2014). Moreover, compared to confirmatory factor analytic methods, IRT is more robust and informative when examining the equivalence of a particular scale (Meade and Lautenschlager 2004).

We first assessed uni-dimensionality using principal component analysis (PCA). We then assessed item properties with a uni-dimensional graded item response model (Samejima 1997). To determine model fit, we used the root mean square error of approximation (RMSEA) and chi-square (X^2). RMSEA chooses optimal parameters estimates to examine how well the unknown model fits the population covariance matrix (Hooper et al. 2008). RMSEA values of 0.06 or lower indicate acceptable model fit. However, due to criticisms of the RMSEA and other fit indices (Hayduk et al. 2007; Squires et al. 2013), we have also reported X^2 values for model fit.

We conducted our analyses using SPSS version 24, R Studio version 3.1.1, and IRTPRO version 2.1 (Cai et al. 2011).

Study 1: Expectations of Computerized Cognitive Training

We evaluated expectations of CCT outcomes on two separate occasions: (i) as part of our initial development of the survey, evaluating 422 participants (305 women; $M = 29.67$ years of age, $SD = 19.76$), and (ii) as a replication in 373 participants (229 women; $M = 46.13$ years of age, $SD = 21.81$). We excluded responses from 153 participants with incomplete or incomparable responses due to their completion of the initial version of the survey, as well as 23 outliers identified based on a Mahalanobis distance greater than 22.46, for a final sample of 619.

Study 2: Expectations of Non-invasive Brain Stimulation

We evaluated expectations of NIBS outcomes in 516 respondents (299 women; $M = 32.22$ years of age, $SD = 17.48$). We excluded 96 participants with inappropriate or incomplete responses, as well as 7 outliers based on a Mahalanobis distance greater than 20.52, for a final sample of 413. We probed expected outcomes for six cognitive domains (i.e., six items): "general cognitive function," "memory," "distractibility," "reasoning ability," "multitasking ability," and "performance in everyday activities"; we dropped the "concentration" item due to a technical error leading to insufficient number of responses.

Because we have also used the EAS in studies of expectation priming (e.g., see Rabipour et al. 2017; Rabipour and Davidson 2015), we also examined the psychometric properties of the EAS when delivered under conditions where participants were primed to report high or low expectations of outcomes.

Results

PCA confirmed the uni-dimensionality of both versions of the EAS, with a ratio of 3.5 and 3.62 between the first and second

Table 1 Composition of participant samples (*n* women) in study 1 and study 2. As previously reported (Rabipour et al. 2017), 88 participants from study 2 did not report age or sex

		UO	FSU	Community	MTurk	Qualtrics
Study 1	YA	143 (112)	50 (28)	213 (153)	94 (28)	–
	MA	–	–	21 (17)	46 (27)	–
	OA	1 (0)	60 (26)	100 (69)	8 (6)	59 (44)
Study 2	YA	171 (141)	–	–	129 (49)	–
	MA	3 (3)	–	–	47 (28)	–
	OA	–	–	–	11 (6)	67 (45)

eigenvalues for the CCT and NIBS scales, respectively. Moreover, all items on each version of the EAS loaded strongly onto a single cognitive domain factor (Table 2). We therefore did not perform confirmatory factor analyses.

Expectations of Computerized Cognitive Training

The hypothesized IRT model for the CCT version of the EAS (EAS-CCT) showed good fit (RMSEA = 0.04; M^2_{749} = 1563.26, $P = 0.0001$). Furthermore, item fit statistics demonstrated that all items fit well with the model (Table 3).

IRT analysis yielded a high reliability of 0.87 for responses on the EAS-CCT. The items with the highest discrimination (i.e., ability to discriminate between high and low expectancy traits) in expectations of CCT were "general cognitive function" (2.34), "memory" (2.27), and "concentration" (2.22); other items had lower discrimination (Table 4). The item information curves show that each item is reliable at each level of the trait, with higher peaks representing higher

Table 2 Principal component analysis for the CCT and NIBS versions of the EAS. Factor loadings suggest all items load strongly onto a single cognitive domain factor

Cognitive domain	λ_1	s.e.
a) CCT		
General cognitive function	0.81	0.04
Memory	0.80	0.04
Concentration	0.79	0.04
Distractibility	0.66	0.05
Reasoning ability	0.69	0.05
Multitasking ability	0.64	0.05
Performance in everyday activities	0.71	0.05
b) NIBS		
General cognitive function	0.75	0.05
Memory	0.85	0.04
Distractibility	0.69	0.06
Reasoning ability	0.83	0.04
Multitasking ability	0.82	0.04
Performance in everyday activities	0.85	0.04

discrimination on the item (Fig. 1). For example, the "memory" item (Fig. 1b) is the most discriminating item with particularly good reliability at the lower end of the expectation spectrum. Conversely, the "multitasking" item demonstrates the lowest discrimination across all levels of expectancy (Fig. 1f). Comparing all the item information curves shows that the items with the lowest discrimination provide the same amount of information across all levels of expectancy. The total information curve shows that the EAS-CCT is more reliable at lower levels of expectancy.

Expectations of Non-invasive Brain Stimulation

The hypothesized model for the NIBS version of the EAS (EAS-NIBS) showed moderately good fit (RMSEA = 0.07; M^2_{534} = 1470.68, $P = 0.0001$). Item fit characteristics demonstrate that all items fit well with the model (Table 5).

IRT analysis yielded a reliability of 0.91 for responses on the EAS-NIBS. The item with the highest discrimination in expectations of NIBS was "performance in everyday activities" (2.72), although most items had high discriminations (Table 6). The item information curves show that each item is reliable at each level of the trait, with higher peaks representing higher discrimination on the item (Fig. 2). For example, the "performance in everyday activities" item (Fig. 2f) discriminates best between respondents of high vs. lower expectancy trait, with particularly good reliability at the lower end of the expectation spectrum. This can be seen both in the

Table 3 Item fit statistics for the EAS-CCT

Cognitive domain	X^2	*d.f.*	Probability
General cognitive function	90.95	67	0.0274
Memory	99.53	61	0.0013
Concentration	106.32	65	0.0009
Distractibility	142.14	80	0.0001
Reasoning ability	108.81	80	0.0178
Multitasking ability	119.67	84	0.0064
Performance in everyday activities	114.08	82	0.0111

Table 4 Item characteristics for the EAS-CCT. Item discrimination and threshold parameters for the EAS-CCT. Items demonstrate good model fit

| Cognitive domain	a	s.e.	b_1	s.e.	b_2	s.e.	b_3	s.e.	b_4	s.e.	b_5	s.e.	b_6	s.e.
General cognitive function	[7]2.34	0.18	− 3.41	0.33	− 2.20	0.14	− 1.56	0.10	− 0.77	0.06	0.48	0.07	2.08	0.14
Memory	[14]2.27	0.18	− 3.46	0.34	− 2.44	0.17	− 1.73	0.11	− 0.95	0.07	0.56	0.07	2.29	0.15
Concentration	[21]2.22	0.17	− 3.46	0.34	− 2.32	0.16	− 1.43	0.09	−0.70	0.06	0.67	0.07	2.31	0.15
Distractibility	[28]1.51	0.12	− 3.44	0.31	− 2.32	0.18	− 1.24	0.10	− 0.23	0.07	1.15	0.11	2.96	0.23
Reasoning ability	[35]1.64	0.13	− 3.07	0.25	− 2.19	0.16	− 1.42	0.11	− 0.37	0.07	0.85	0.09	2.72	0.20
Multitasking ability	[42]1.42	0.11	− 3.31	0.29	− 2.24	0.17	− 1.40	0.11	− 0.39	0.07	0.85	0.10	2.37	0.19
Performance in everyday activities	[49]1.72	0.13	− 3.13	0.26	− 2.10	0.15	− 1.43	0.10	− 0.38	0.07	0.98	0.09	2.52	0.18

table as well as in the height of the peaks in Fig. 2. Figure 2f demonstrates that the "memory" item discriminates well across all levels of the expectancy trait, with particularly high reliability at lower levels of the trait. Conversely, the "distractibility" item demonstrates the lowest discrimination across all levels of expectancy (Fig. 2c). Comparing all the item information curves shows that the items with the lowest discrimination provide the same amount of information across all levels of expectancy. The total information curve shows that the EAS-NIBS is more reliable at lower levels of expectancy.

As predicted, we found substantial differences in the psychometric properties of both versions of the EAS under conditions of expectation priming. Specifically, despite good reliability under priming conditions, both versions of the EAS demonstrated poor fit with the hypothesized model under high (EAS-CCT: RMSEA = 0.31; $M^2_{749} = 45,088.54$, $P = 0.0001$;

EAS-NIBS: RMSEA = 0.24; $M^2_{534} = 13,051.28$, $P = 0.0001$) and low (EAS-CCT: RMSEA = 0.26; $M^2_{749} = 32,120.61$, $P = 0.0001$; EAS-NIBS: RMSEA = 0.16; $M^2_{534} = 6397.45$, $P = 0.0001$) priming conditions.

Discussion

Expectations may influence the outcomes of cognitive enhancement interventions, but remain unaccounted for in the majority of trials. Here, we demonstrate the reliability of a new tool, the EAS, designed specifically to measure expectations of outcomes in the context of cognitive enhancement interventions. Using IRT, we have found that the EAS has internal consistency and can serve as a reliable measure of participant expectations.
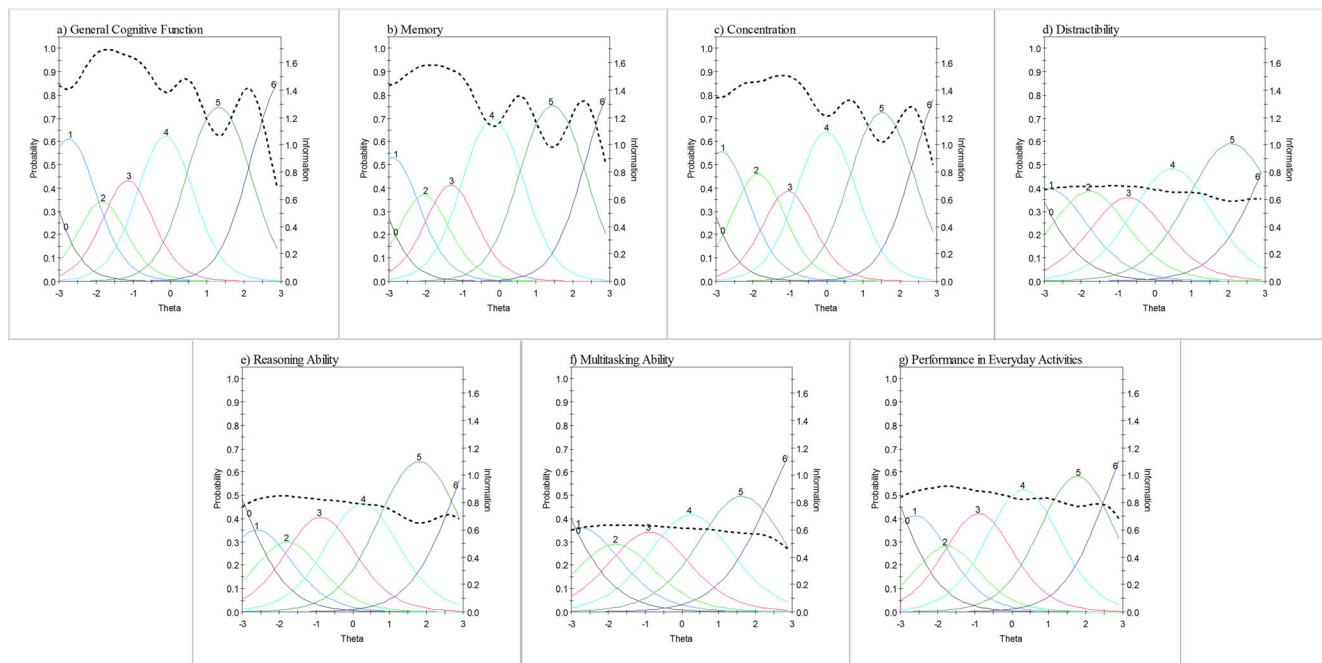
Fig. 1 Category characteristic and item information curves (dashed line) for the EAS-CCT. Results suggest the CCT version of the EAS discriminates responses well at all levels of the trait, for all included subscales (**a**–**g**). Category characteristic curves show the best discriminability in the "general cognitive function" subscale (**a**)

Table 5 Item fit characteristics of the EAS-NIBS

Cognitive domain	X^2	d.f.	Probability
General cognitive function	119.50	66	0.0001
Memory	121.87	62	0.0001
Distractibility	126.79	79	0.0005
Reasoning ability	145.33	67	0.0001
Multitasking ability	97.51	66	0.0071
Performance in everyday activities	106.66	61	0.0003

Our results further show good validity based on the internal structure of the EAS across all subscales, for both CCT and NIBS. Response patterns are comparable across multiple samples of young and older adults, and suggest that older adults may have higher expectations of CCT compared to young (Rabipour et al. 2017; Rabipour and Davidson 2015). Moreover, although current analyses of the EAS cannot determine the extent to which responses are truly reflective of expectations, the EAS appears sensitive to changes in reported expectations, relative to baseline, in response to information indicating high or low intervention effectiveness. Because no distinct expectation scale currently exists, we cannot perform an analysis of external validation. However, we would argue that the EAS has high face validity.

Of course, the reliability of the EAS does not necessarily indicate how participants might have interpreted the purpose of the study or the reasoning behind their expectation ratings. Similarly, our results do not provide any insight into how participants interpreted each cognitive domain (i.e., item) in the EAS; notably, the different discriminability of the "concentration" and "distractibility" items, intended to be mirror opposites, suggests that participants may have perceived them as diverging. However, these potential considerations hardly undermine the utility of the tool. The EAS aims to capture how the average person—in particular, non-experts interested in cognitive enhancement interventions but naïve to the scientific context of CCT and NIBS—would respond when reading freely available information supporting or criticizing CCT or NIBS programs (e.g., through advertisements, consumer reviews, or personal anecdotes). Such information,

which often fuels people's decision whether or not to pursue a particular program, formed the basis of our expectation priming messages (Rabipour and Davidson 2015). Determining how people would respond in this type of hypothetical scenario represents an important first step to understanding people's motivations to pursue such interventions when presented with the opportunity.

Limitations

A potential limitation is our inclusion of seven items in study 1 and only six in study 2. In addition, differential recruitment (i.e., lack of recruitment from FSU and from the Ottawa community for Study 2), as well as the inclusion of 50 OA recruited in the context of a brain training intervention in study 1, could potentially explain result patterns across our studies.

Conclusion

The EAS can serve as a reliable and valid tool to assess the face validity of cognitive enhancement interventions, while accounting for expectations of outcomes. Using the EAS may help interpret the results of cognitive enhancement trials and determine the effectiveness and appropriateness of individual approaches in different populations. Our data may serve as a reference point to create a standardized expectation scale, with external validation and population norms.

Future studies may refine or customize the structure of the EAS as relevant, based on individual program structure, and use alternative approaches (e.g., involving factor analytic methods) for validating the tool. One useful next step would be to compare intervention outcomes in people who receive the EAS before beginning a study vs. those who do not, as suggested by our reviewers. Alternatively, allowing a certain period of time to pass between administration of the EAS and the intervention might be a good way to avoid having the EAS influence outcomes in an unexpected or artificial way. In addition, as suggested by a reviewer, future research could examine the psychometric properties of the EAS in situations where participants are given a basis for discriminating among

Table 6 Item characteristics of the EAS-NIBS. Parameter estimates and discrimination indices for the EAS-NIBS. Items demonstrate good model fit

Cognitive domain	a	s.e.	b_1	s.e.	b_2	s.e.	b_3	s.e.	b_4	s.e.	b_5	s.e.	b_6	s.e.
General cognitive function	[7]1.93	0.19	−2.99	0.27	−1.71	0.15	−0.91	0.09	0.05	0.07	1.30	0.12	2.87	0.28
Memory	[14]2.69	0.26	−2.21	0.16	−1.18	0.09	−0.52	0.06	0.28	0.06	1.33	0.11	2.36	0.21
Distractibility	[21]1.63	0.15	−2.56	0.22	−1.49	0.13	−0.67	0.09	0.28	0.08	1.55	0.15	2.82	0.28
Reasoning ability	[28]2.57	0.24	−1.72	0.12	−0.85	0.08	−0.28	0.06	0.51	0.07	1.43	0.12	2.40	0.22
Multitasking ability	[35]2.44	0.22	−1.66	0.12	−0.82	0.07	−0.30	0.06	0.45	0.07	1.32	0.11	2.40	0.21
Performance in everyday activities	[42]2.72	0.25	−1.95	0.14	−1.00	0.08	−0.54	0.06	0.19	0.06	1.11	0.09	2.34	0.20
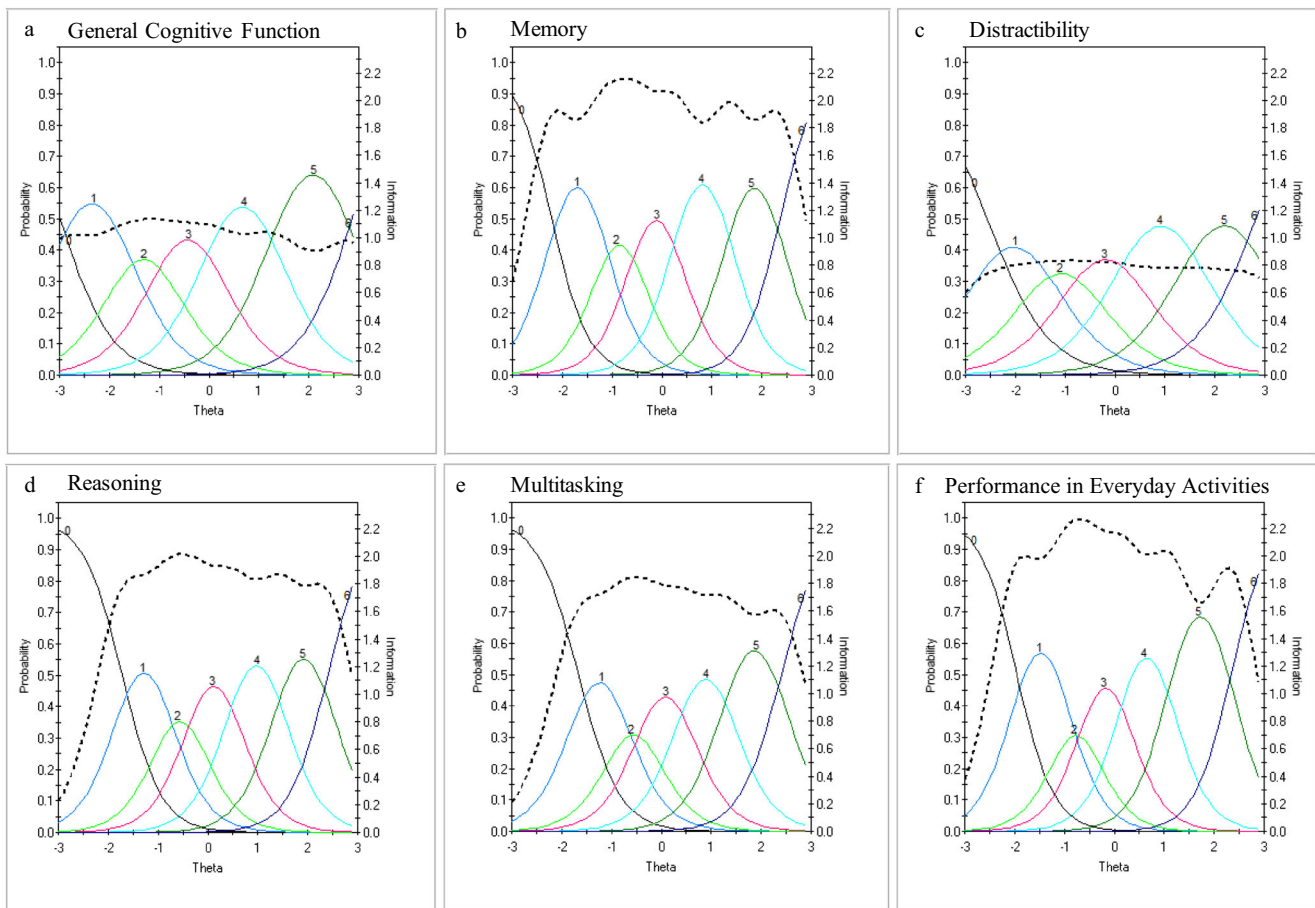
Fig. 2 Category characteristic curves and item information curves (dashed line) for the EAS-NIBS. Results suggest the NIBS version of the EAS discriminates responses well at all levels of the trait, for all included subscales (**a**–**f**). Category characteristic curves suggest the best discriminability in the "performance in everyday activities" subscale (**f**)

the likely outcomes (e.g., providing information about the implicated brain regions and their associated functions).

Assessing expectations with the EAS before, during, and after cognitive enhancement interventions will likely prove useful for explaining intervention outcomes, identifying differences between responders and non-responders, and better tailoring interventions to target users.

References

Boot, W. R., & Kramer, A. F. (2014). The brain-games conundrum: does cognitive training really sharpen the mind?. Retrieved from http://www.dana.org/Cerebrum/2014/The_Brain-Games_Conundrum__Does_Cognitive_Training_Really_Sharpen_the_Mind_/-sthash.iPYa8y4b.dpuf.

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science, 8*(4), 445–454. https://doi.org/10.1177/1745691613491271.

Cai, L., Thissen, D., & du Toit, S. H. C. (Producer). (2011). IRTPRO for Windows [Computer software].

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Psychology Press.

Foroughi, C. K., Monfort, S. S., Paczynski, M., McKnight, P. E., & Greenwood, P. M. (2016). Placebo effects in cognitive training. *Proceedings of the National Academy of Sciences of the United States of America, 113*(27), 7470–7474. https://doi.org/10.1073/pnas.1601243113.

Foster, G. C., Min, H., & Zickar, M. J. (2017). Review of item response theory practices in organizational research: lessons learned and paths forward. *Organizational Research Methods, 20*(3), 465–486. https://doi.org/10.1177/1094428116689708.

Hayduk, L., Cummings, G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! Testing! One, two, three—testing the theory in structural equation models! *Personality and Individual Differences, 42*(5), 841–850. https://doi.org/10.1016/j.paid.2006.10.001.

Hays, R. D., Morales, L. S., & Reise, S. P. (2000). Item response theory and health outcomes measurement in the 21st century. *Medical Care, 38*(9), 28–42 Retrieved from <go to ISI>://WOS:000089033000005.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53–60.

Kim, S., & Feldt, L. S. (2010). The estimation of the IRT reliability coefficient and its lower and upper bounds, with comparisons to CTT reliability statistics. *Asia Pacific Education Review, 11*(2), 179–188. https://doi.org/10.1007/s12564-009-9062-8.

Meade, A. W., & Lautenschlager, G. J. (2004). A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organizational Research Methods, 7*(4), 361–388. https://doi.org/10.1177/1094428104268027.

Rabipour, S., & Davidson, P. S. R. (2015). Do you believe in brain training? A questionnaire about expectations of computerised cognitive training. *Behavioural Brain Research, 295*, 64–70. https://doi.org/10.1016/j.bbr.2015.01.002.

Rabipour, S., Andringa, R., Boot, W. R., & Davidson, P. S. R. (2017). What do people expect of cognitive enhancement? *Journal of Cognitive Enhancement, 1–8.* https://doi.org/10.1007/s41465-017-0050-3.

Samejima, F. (1997). Graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 85–100). Basel: Springer International Publishing AG.

SharpBrains. (2013). Executive summary: infographic on the digital brain health market 2012–2020. Retrieved from http://www.sharpbrains.com/executive-summary/.

SharpBrains. (2016). The digital brain health market 2012–2020: web-based, mobile and biometrics-based technology to assess, monitor and enhance cognition and brain functioning. Retrieved from http://sharpbrains.com/market-report/.

Shu, L. H., & Schwarz, R. D. (2014). IRT-estimated reliability for tests containing mixed item formats. *Journal of Educational Measurement, 51*(2), 163–177. https://doi.org/10.1111/jedm.12040.

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do "brain-training" programs work? *Psychological Science in the Public Interest, 17*(3), 103–186. https://doi.org/10.1177/1529100616661983.

Souders, D. J., Boot, W. R., Blocker, K., Vitale, T., Roque, N. A., & Charness, N. (2017). Evidence for narrow transfer after short-term cognitive training in older adults. *Frontiers in Aging Neuroscience, 9.* https://doi.org/10.3389/fnagi.2017.00041.

Squires, J. E., Hayduk, L., Hutchinson, A. M., Cranley, L. A., Gierl, M., Cummings, G. G., …, Estabrooks, C. A. (2013). A protocol for advanced psychometric assessment of surveys. *Nursing Research and Practice, 2013*, 156782. https://doi.org/10.1155/2013/156782.