HHS Public Access
Author manuscript
Linguist Approaches Biling. Author manuscript; available in PMC 2017 May 10.

Published in final edited form as:
Linguist Approaches Biling. 2016 ; 6(5): 535–548. doi:10.1075/lab.15041.fri.

Research on Individual Differences in Executive Functions: Implications for the Bilingual Advantage Hypothesis

Naomi P. Friedman
University of Colorado at Boulder

Abstract

Executive functions (EFs), such as response inhibition, interference control, and set shifting, are general-purpose control mechanisms that enable individuals to regulate their thoughts and behaviors. Because bilingual individuals use EF-like processes during language control, researchers have become interested in the hypothesis that this use might train EFs, resulting in better performance on non-linguistic EF tasks. Although this bilingual advantage hypothesis seems straightforward to test, it involves a number of important decisions in terms of how to assess bilingualism and EFs. In this article, I focus on the complexity of measuring EFs, drawing on individual differences research (conducted with participants not selected for bilingualism). Specifically, I discuss issues related to (1) the measurement of EFs (particularly the effects of task impurity and unreliability) and (2) the multicomponent nature of EFs. Within each of these topics, I elaborate on consequences for research on bilingual advantages and provide some recommendations.

Keywords

Executive control; executive functioning; multilingualism; latent variables

Executive functions (EFs) are general-purpose control mechanisms that enable people to regulate their thoughts and behaviors to align with their goals. For example, stopping an automatic response, ignoring irrelevant information, and switching between multiple tasks are all considered EFs (Diamond, 2013; Miyake & Friedman, 2012). Because bilingual individuals must use EF-like processes everyday (e.g., selecting one language in the correct context, ignoring irrelevant information about the same concepts in the unselected language, switching between languages), researchers have become interested in the hypothesis that this use might train EFs, resulting in better performance on non-linguistic EF tasks (Bialystok, Craik, & Luk, 2012). This *bilingual advantage* hypothesis seems straightforward to test: Obtain a sample of bilinguals and appropriately matched controls and test whether they differ on a measure of EF, using an analysis model like the one depicted in Figure 1a. Yet within that simple design are a number of important decisions. What kind of bilinguals should be tested? What aspect of EF should be assessed? How should EF(s) be measured?

Correspondence concerning this article should be addressed to Naomi Friedman, Institute for Behavioral Genetics, 447 UCB, University of Colorado, Boulder, CO 80309. naomi.friedman@colorado.edu.
Naomi P. Friedman, Department of Psychology and Neuroscience and Institute for Behavioral Genetics, University of Colorado Boulder.

Each of these decisions can have a large influence on the study outcome and conclusions that can be drawn. A number of papers have discussed the complexity of measuring bilingualism and obtaining appropriate control groups; in this article, I focus on the complexity of measuring EFs, drawing on individual differences research. Specifically, I discuss issues related to (1) the measurement of EFs and (2) the multicomponent nature of EFs. Within each of these topics, I elaborate on consequences for research on the bilingual advantage hypothesis and provide some recommendations.

1. Measuring EFs

1.1. Individual tasks

The model in Figure 1a depicts a test of the bilingual advantage hypothesis on one observed EF task. Some of the most frequently examined tasks in this context are ones that require stopping a dominant response and/or resolving interference, like the Simon or Eriksen flanker tasks (see Valian, 2015, for descriptions and links to demonstrations of these and other frequently used EF tasks). Yet, as a number of bilingualism researchers have noted (e.g., Paap & Greenberg, 2013; Valian, 2015), the use of a single EF measure can be problematic, because inter-correlations among EF tasks taken to tap the same cognitive processes are often quite low. For example, Friedman and Miyake (2004) reported correlations ranging from .11 to .18 for three tasks designed to measure individual differences in interference control (including the Eriksen flanker task) in a sample of 220 undergraduate students, a result that was entirely consistent with the prior literature.

Such low correlations are problematic because they suggest that no one task is capturing much variance related to the EF of interest. If two tasks require the same cognitive process, then scores on both tasks should be influenced by individual differences in that ability, and their covariance should reflect the extent to which that ability influences both tasks. EF tasks have low inter-correlations for several reasons, perhaps the most important of which is the *task impurity* problem (Miyake et al., 2000). By definition, EFs are higher-level control processes that operate on lower-level processes, so the lower-level processes must be included in the task. If individual differences in those non-EF processes also influence performance, then scores on that task will not be pure measures of the EF of interest, but also contain non-EF variance. For example, the Stroop task involves resisting a prepotent word reading response in favor of a color naming response, but scores may also be influenced by individual differences in reading ability, color vision, and general speed. Thus, a poor score on this task may reflect low EF, but also may reflect individual differences in these other processes.

In the context of testing the bilingual advantage hypothesis, task impurity can lead to both type II and type I errors. First, if bilinguals do have an advantage on an EF tapped impurely by the chosen task, then power will be lower to detect that effect, because only part of the task variance reflects individual differences in that EF. If sample size is not adequate (see section 1.2 for a discussion of how much effects can be attenuated and section 1.3 for an example calculation of sample size requirements), the result may be a type II error. Second, if bilinguals do not have an EF advantage, they may nevertheless show better performance on a task because they excel at the non-EF aspects of the task. To the extent that this benefit holds across samples, this result may be replicated across many labs. Although such a difference may well be interesting, it may be considered a type I error if it interpreted as evidence for an EF advantage.

Many EF tasks include control conditions to try to remove the influence of non-EF processes (e.g., scores on conflict tasks are typically difference scores between reaction times for incongruent stimuli vs. congruent or neutral stimuli). The use of difference scores can help with the task impurity problem but may not completely rectify it (e.g., if the influence of the non-EF processes is not linear), and can exacerbate another problem with some EF tasks: low reliability. EF tasks in general (not just those based on difference scores) often have somewhat low reliability, which likely has several sources (such as variable strategies that participants use within or across sessions, or decreases in the extent to which EFs are needed to complete tasks once they are no longer novel; Rabbitt, 1997) and difference scores tend to show lower reliability than the components that go into those differences (Johns, 1981). Because reliability puts an upper limit on how well a task will correlate with other constructs of interest, the consequence of this problem is also lower power for examining the bilingual advantage hypothesis.

1.2. Latent variables

To reduce the influence of task impurity and low reliability, many researchers measure EFs with latent variables (illustrated in Figure 1b). A latent variable is a hypothetical or unobservable variable for which true scores cannot be directly measured, but can be inferred based on performance in tasks that it influences. Thus, it is estimated as an unobserved variable (shown with an ellipse in a diagram) that influences performance (via single-headed arrows with factor loadings λ_i) across multiple observed tasks (shown as rectangles). The variance in each task can be partitioned into that due to the latent variable and a unique component (ε_i, which includes reliable variance that is not due to the latent variable as well as measurement error). If the tasks that define the latent variable are selected to differ in their non-EF demands (i.e., by considering what lower-level cognitive processes may also lead to individual differences in performance, either in theory or as informed by prior literature), then the common variance captured by the latent variable will be a purer, more reliable measure of the construct of interest (Bollen, 1989).

Regardless of whether one actually uses multiple tasks and estimates a latent variable, the properties of the latent variable have important implications for tests of the bilingual advantage hypothesis. That is, even when only one task is used, the measurement properties of that task, as revealed by latent variable analyses in other studies, are the same. So, if only a portion of the variance in a task reflects the underlying EF of interest, then only that portion of the task variance can be reasonably expected to vary with bilingualism according to the bilingual advantage hypothesis; the rest of the non-EF variance would essentially be error in this context.

One can predict how much variance in a task can be reasonably expected to relate to bilingualism by examining correlations of that task with other tasks thought to tap the same EF, because the standardized loadings on a latent variable are determined by the inter-correlations of the tasks. A loading can be interpreted as a regression of the task score on the latent variable, where the squared loading equals the variance in the task explained by the latent variable. For example, if the three tasks loading on the EF factor in Figure 1b correlate .25, then their standardized loadings (λ_i) would be .50, and we would say that the latent variable explains 25% of the variance in each task. The effect of a predictor, such as bilingualism, on any one of those tasks would be the product of the effect on the latent variable (β_1; i.e., the true effect) and that task's loading (λ_i). So, for example, if bilinguals were 0.5 SD (β_1) better at interference control than monolinguals, and each task loads .50 (λ_i) on the interference control latent variable, then we should expect a 0.5*.50=0.25 SD difference between bilinguals and monolinguals on each task. This example shows how the effect sizes one can observe with an individual EF task will be attenuated from the true effect (in this case, from a medium effect size to a small one) by task impurity and measurement error. In the case of tasks like the ones used to assess interference control ability, correlations tend to be lower. For example, Friedman and Miyake (2004) found loadings from 0.32 to 0.42 on an interference control latent variable that included the Eriksen flanker task.

This situation is unfortunately the typical scenario in research on the bilingual advantage. One can recover the true effect size by using latent variables, but there are also costs in doing so. Completing multiple tasks per construct increases the burden for the participants, and latent variable analysis requires large sample sizes to obtain good estimates (though simple models can be successful with fewer subjects, a typical recommendation is at least 200; Kline, 2011). With small samples, one may approximate a latent variable by averaging multiple measures, but this average will be an imperfect estimate of the underlying construct. Ideally, we would find measures that have higher loadings on EF latent variables, but that would require creating and validating new tasks. Moreover, to the extent that any EF task must be impure, there will always be a need for examining multiple measures.

1.3. Recommendations

These measurement properties of EFs lead to the following recommendations. First, if one is using individual EF tasks, one should have large enough sample sizes for adequate power to detect realistic effect sizes, taking into account the attenuation of true effects found with individual EF tasks. For example, suppose one wanted to detect a medium-sized effect (0.5 SD) of bilingualism on interference control with a flanker task. Given an approximate loading of .40 of that task on an interference control latent variable (Friedman & Miyake, 2004), one might expect an observed effect size of .5*.40 = .20, which would require a total sample size of approximately 800 to detect with 80% power.

Most studies examining individual tasks have used much smaller sample sizes (but see Gathercole, 2014, for sample sizes up to 650). Thus, meta-analyses will likely provide the best evidence for or against the bilingual advantage hypothesis. Although a recent meta-analysis (de Bruin, Treccani, & Sala, 2015) concluded that there was a significant effect, it also found evidence for publication bias, which means that their .30 SD effect size is likely an overestimate. Such results raise the possibility that the significant effects that have been found with small sample sizes may be false positives (Button et al., 2013; see Paap, Johnson,

& Sawi, 2015, for an in-depth discussion of this and other possibilities with regard to the bilingual advantage literature).

Second, to ensure that effects are not due to task impurity, one should use multiple measures of the targeted EF and evaluate whether patterns are consistent across these EF tasks. If, for example, bilinguals show an advantage on one measure of interference control but not others, then it is possible that the observed effect is due to an advantage in the non-EF variance in that one task, rather than an advantage in the EF (see von Bastian, Souza, & Gade, 2015, for an example).

Of course, increasing the number of dependent measures increases the number of statistical tests and hence the possibility of a type I error. Testing hypotheses with aggregated measures (such as a *z*-composite or, if sample size permits, a latent variable) will reduce the number of tests at the same time that it reduces the influence of task impurity and unreliability. However, the usefulness of this aggregate depends on the tasks that go into it, so care should be taken to select reliable and valid measures that tap the same EF; ideally, task selection should be guided by current theory regarding the structure of EFs, discussed in the next section.

2. Multi-Component Nature of EFs

2.1. Unity and diversity

As mentioned earlier, the term EF has been used to describe a broad range of cognitive control processes, including suppressing prepotent responses, resisting attentional distraction, resisting proactive interference, switching between task sets, updating working memory, verbal and spatial working memory capacity, dual tasking, planning, monitoring, and fluency. A large body of research suggests that these EFs are a family of functions, rather than a unitary construct. Although various EFs seem to share something in common, they also have unique variances, a pattern described as "unity and diversity" (e.g., Miyake et al., 2000).

For example, Miyake et al. (2000) examined the correlations among three commonly studied EFs — prepotent response inhibition, working memory updating, and task shifting — at the level of latent variables. To do so, they measured undergraduate students on a battery of tasks selected to tap particular EFs while differing in the lower-level processes on which those EFs operated. Inhibiting is the ability to stop a dominant or automatic response (sometimes in order to make an alternative response), which they measured with tasks that required stopping reflexive eye movements, stopping a well-practiced categorization response, and stopping a word-naming response. Updating is the ability to monitor the environment for relevant information and continuously replace no-longer relevant information in working memory with new relevant information, which they measured with tasks that required updating words from different categories, letters, or auditory tones. Finally, Shifting is the ability to rapidly switch between two task sets, which they measured with switch costs (the time to switch to a different subtask minus the time to repeat the same subtask) in tasks that required switching between adding and subtracting, categorizing numbers and letters, or attending to local vs. global features of an image. Miyake et al.

found that these three EFs were significantly correlated (correlations ranged from .42 to .63) but separable (i.e., no two factors could be collapsed) at the level of latent variables. This basic pattern has been replicated numerous times in independent samples with different tasks and varied age ranges (see Miyake & Friedman, 2012, for a review).

Usually, EF latent variable models include multiple correlated factors (e.g., Inhibiting, Updating, and Shifting latent variables allowed to correlate; see Figure 2a). Miyake and Friedman (2012) discussed the benefits of an alternative bifactor parameterization (Figure 2b) in which a common factor (Common EF) predicts all tasks, and orthogonal specific factors (Updating-Specific and Shifting-Specific factors) capture additional variance not captured by the Common EF factor. Fit-wise, such a model is not that different from the correlated factors model, but it can be useful conceptually for examining the unity and diversity components directly. That is, in the correlated factors model, unity and diversity are captured by the inter-factor correlations, whereas in this bifactor model, they are captured directly by the latent variables.

To see why this model might be useful, consider this example. Suppose one found a bilingual advantage in Inhibiting, Updating, and Shifting abilities (either at the level of latent variables or individual tasks selected to tap each of these constructs). Because these three EFs are correlated, it would be hard to say whether bilinguals were better at several processes unique to particular EFs (i.e., diversity components), better at some process common to all three (i.e., the unity component), or some combination of these possibilities. However, using the bifactor model, one could directly test whether bilingualism influenced the common factor as well as the unique factors.

2.2. No inhibition-specific ability

One particularly interesting finding that has been replicated in several independent datasets (Miyake & Friedman, 2012) is that there is no evidence for an inhibiting-specific factor. That is, once the Common EF factor is in the model, there are no remaining correlations among the inhibition tasks to create an additional factor; the Common EF factor captures all the variance in response inhibition ability. Because inhibition is such a key construct in the concept of executive control, this finding has been of particular theoretical interest.

One interpretation of this result is that Common EF is really inhibition ability (e.g., Valian, 2015). This viewpoint posits that EF tasks correlate because they all require some sort of inhibition. However, this hypothesis assumes a very broad definition of inhibition that lumps together various conceptually and empirically distinct abilities (such as inhibition of responses vs. no-longer-relevant memory contents). For example, Friedman and Miyake (2004) found that cognitive processes that are all called "inhibition" (response inhibition, resisting distractor interference, and resisting proactive interference) were not all highly correlated at the latent variable level: Although the response inhibition and resistance to distractor interference factors could be collapsed into a single factor, that factor was not significantly related to resistance to proactive interference, suggesting that so-called inhibition tasks do not all tap the same ability.

An alternative interpretation of the absence of an inhibiting-specific factor is that individual differences in response inhibition are determined primarily by a more general process that is common to EF tasks, namely actively maintaining task goals in the face of interference and using these goals to bias lower level processing (Miyake & Friedman, 2012). Strong goal representations increase activation for goal-relevant information, and decrease activation for irrelevant information through competition (via local lateral inhibition). Thus, inhibition (of distractors, responses, memory representations, etc.) emerges from neural competition rather than being a key function of the frontal lobe areas implicated in EF tasks (Munakata et al., 2011). This proposal is consistent with numerous existing theories of EF and frontal lobe functioning (e.g., Miller & Cohen, 2001), as well as goal-directed views of inhibitory control and conflict resolution (e.g., Banich & Depue, 2015; Munakata et al., 2011).

The mechanisms underlying inhibitory control (particularly response inhibition) are still debated (see Aron, Robbins, & Poldrack, 2014; Banich & Depue, 2015). The resolution of this debate will be important for the study of bilingual advantage (and EF training more generally), because these studies posit that advantages arise from particular mechanisms being more practiced. If researchers focus only on top-down inhibition as a mechanism of interest, they may lose sight of other mechanisms that could explain performance and training benefits. For example, building on the idea that individual differences in response inhibition may relate to goal-related processes rather than inhibition, Chevalier, Chatham, and Munakata (2014) found that having children practice context monitoring for goal-relevant signals (with no stopping) was more beneficial for later stopping performance than practicing actually stopping. They interpreted this finding as evidence that such monitoring processes are "critical to developing inhibitory control and suggest promising new directions for interventions" (p. 964).

2.3. Recommendations

As the previous paragraphs make clear, there is good evidence for the unity and diversity of EFs, and the mechanisms underlying this structure are being actively researched and debated. This state of the field leads to the following recommendations.

First, given the fractionation of EFs, it is important to consider what construct is of interest, and how it can best be measured, when designing a study. For example, given evidence that not all forms of "inhibition" are closely related (e.g., Friedman & Miyake, 2004), one might include multiple measures of one kind of inhibition, as opposed to measures of dissociable kinds (unless testing for dissociations is a goal). von Bastian et al. (2015) provides an excellent example of a such a design that they used to test multiple forms of the bilingual advantage hypothesis (this study also provides examples of task adaptations, given that some popular EF tasks have linguistic requirements that make them inappropriate for use with bilinguals). This comprehensive study yielded no consistent evidence for an advantage in inhibitory control, conflict monitoring, shifting, or general cognitive ability, although it is possible, given the earlier discussion of power, that they were underpowered even with their sample of 118 participants.

Second, it is important to consider the unity of EFs as well as the diversity. In particular, finding a bilingual advantage in one EF such as set shifting may not mean that the key benefit is in Shifting, as Shifting is correlated with other EFs. The benefit could thus be more general, as discussed by Kroll and Bialystok (2013). To evaluate such a possibility, one would need to examine multiple EFs, optimally within the same study to evaluate whether benefits occur within unity or diversity components, or both.

More generally, as many researchers have already recognized, it may be useful to consider multiple explanations for variability in performance. If performance on a so-called inhibition task is not really driven by top-down inhibition, the interpretation of a bilingual advantage on that task may change. For example, Bialystok et al. (2012) discussed the idea that the bilingual advantage in conflict tasks may reflect improvement in selecting goal-relevant information, rather than suppressing irrelevant information. Other researchers have also considered alternatives to inhibitory views (e.g., Colzato et al., 2008; Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009; Hilchey & Klein, 2011), although these alternatives are usually considered in terms of different kinds of measures (i.e., conflict tasks vs. response inhibition tasks), rather than an evaluation of the possibility that individual differences in classic response inhibition tasks like stop signal may not reflect top-down inhibition.

3. Summary

In this article, I have discussed the measurement and conceptualization of EFs, focusing on implications for tests of the bilingual advantage hypothesis, and made several recommendations, summarized here:

1. Obtain sample sizes large enough to detect reasonable effect sizes, considering attenuation due to task impurity and unreliability of many EF tasks. Such sample sizes are likely larger than those typically used in studies of the bilingual advantage.
2. Include multiple measures of the target EF(s) and evaluate consistency across these measures (or at the level of latent variables) to ascertain whether significant effects reflect advantages in the underlying EF(s) as opposed to task-specific processes.
3. Incorporate the literature on the structure of EFs in interpreting findings. Tasks may show different patterns because they tap dissociable EFs (i.e., diversity); conversely, tasks described as tapping a particular EF may show effects because they also tap a more general process (i.e., unity).
4. Consider alternative mechanisms that could explain individual differences in a particular EF component. Such alternatives could lead to different conclusions or new insights with respect to language control.

The bilingual advantage hypothesis is a hypothesis about training — that practice with language control transfers to general cognitive control benefits. Controversies within the literature on cognitive training (e.g., Shipstead, Redick, & Engle, 2012) notwithstanding, a central question is whether the control processes used in the novel EF tasks used to assess bilingual advantages are actually the same as those exercised by the language control practiced by highly proficient bilinguals. If so, then it is plausible that everyday language control may indeed train EFs. But if not, EF abilities may not provide insight into the bilingual brain. My hope is that the discussion and recommendations here, reflecting lessons learned from research outside the area of bilingualism, may be useful for resolving inconsistencies in the bilingual advantage literature and answering this question.

Acknowledgments

This research was supported by a grant from the National Institute of Health (MH063207).

References

Aron AR, Robbins TW, Poldrack RA. Inhibition and the right inferior frontal cortex: One decade on. Trends in Cognitive Sciences. 2014; 18:177–185. [PubMed: 24440116]

Banich MT, Depue BE. Recent advances in understanding neural systems that support inhibitory control. Current Opinion in Behavioral Sciences. 2015; 1:17–22.

Bialystok E, Craik FIM, Luk G. Bilingualism: Consequences for mind and brain. Trends in Cognitive Sciences. 2012; 16:240–250. [PubMed: 22464592]

Bollen, KA. Structural equations with latent variables. New York: Wiley; 1989.

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR. Power failure: Why small sample size undermines the reliability of neuroscience. Nature Reviews Neuroscience. 2013; 14:365–376.

Chevalier N, Chatham CH, Munakata Y. The practice of going helps children to stop: The importance of context monitoring in inhibitory control. Journal of Experimental Psychology: General. 2014; 143:959–965. [PubMed: 24512561]

Colzato LS, Bajo MT, van den Wildenberg W, Paolieri D, Nieuwenhuis S, La Heij W, Hommel B. How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2008; 34:302–312.

Costa A, Hernández M, Costa-Faidella J, Sebastián-Gallés N. On the bilingual advantage in conflict processing: Now you see it, now you don't. Cognition. 2009; 113:135–149. [PubMed: 19729156]

de Bruin A, Treccani B, Sala Della S. Cognitive advantage in bilingualism: An example of publication bias? Psychological Science. 2015; 26:99–107. [PubMed: 25475825]

Diamond A. Executive functions. Annual Review of Psychology. 2013; 64:135–168.

Friedman NP, Miyake A. The relations among inhibition and interference control functions: A latent-variable analysis. Journal of Experimental Psychology: General. 2004; 133:101–135. [PubMed: 14979754]

Friedman NP, Miyake A, Robinson JL, Hewitt JK. Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. Developmental Psychology. 2011; 47:1410–1430. [PubMed: 21668099]

Gathercole VCM. Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. Frontiers in Psychology. 2014; 5:11. [PubMed: 24550853]

Hilchey MD, Klein RM. Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. Psychonomic Bulletin & Review. 2011; 18:625–658. [PubMed: 21674283]

Johns G. Difference score measures of organizational behavior variables: A critique. Organizational Behavior and Human Performance. 1981; 27:443–463.

Kline, RB. Principles and practice of structural equation modeling. New York: Guilford Press; 2011.

Kroll JF, Bialystok E. Understanding the consequences of bilingualism for language processing and cognition. Journal of Cognitive Psychology. 2013; 25:497–514.

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annual Review of Neuroscience. 2001; 24:167–202.

Miyake A, Friedman NP. The nature and organization of individual differences in executive functions: Four general conclusions. Current Directions in Psychological Science. 2012; 21:8–14. [PubMed: 22773897]

Miyake A, Friedman NP, Emerson MJ, Witzki AH, Howerter A, Wager TD. The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. Cognitive Psychology. 2000; 41:49–100. [PubMed: 10945922]

Munakata Y, Herd SA, Chatham CH, Depue BE, Banich MT, O'Reilly RC. A unified framework for inhibitory control. Trends in Cognitive Sciences. 2011; 15:453–459. [PubMed: 21889391]

Paap KR, Greenberg ZI. There is no coherent evidence for a bilingual advantage in executive processing. Cognitive Psychology. 2013; 66:232–258. [PubMed: 23370226]

Paap KR, Johnson HA, Sawi O. Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. Cortex. 2015; 69:265–278. [PubMed: 26048659]

Rabbitt, P. Introduction: Methodologies and models in the study of executive function. In: Rabbitt, P., editor. Methodology of frontal and executive function. Hove, UK: Psychology Press; 1997. p. 1-38.

Shipstead Z, Redick TS, Engle RW. Is working memory training effective? Psychological Bulletin. 2012; 138:628–654. [PubMed: 22409508]

von Bastian, CC.; Souza, AS.; Gade, M. No evidence for bilingual cognitive advantages: A test of four hypotheses. Journal of Experimental Psychology: General. 2015. Advance online publication. http://dx.doi.org/10.1037/xge0000120

Valian V. Bilingualism and cognition. Bilingualism: Language and Cognition. 2015; 18:3–24.

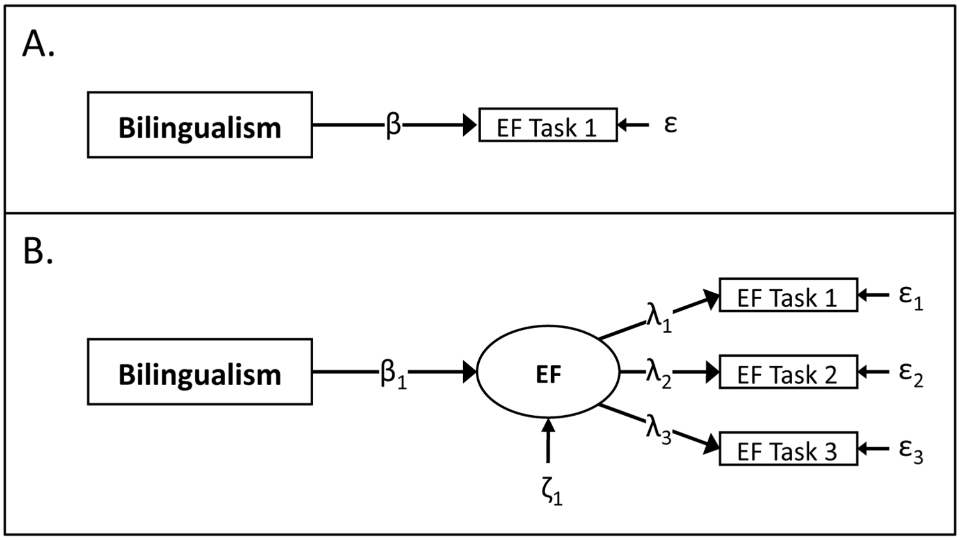

Figure 1.

Models of the bilingual advantage hypothesis with a single executive function (EF) task (panel A) vs. an EF latent variable (panel B). In panel A, β is the direct effect of bilingualism on the EF measure. In panel B, bilingualism influences the EF tasks through the latent EF variable (path β_1). Squaring the loadings of the tasks on the latent variable (λ_i) provides an estimate of the variance in each task attributable to the EF; the remaining variance (i.e., task impurity and unreliability) is captured by the ε_i. In this way, latent variables separate true EF variance from task impurity and unreliability, allowing for an estimate of the true effect on the underlying EF. If these models were estimated with the same data, β from panel A would equal β_1*λ 1 from panel B. Thus, when individual tasks are used as in panel A, the true effect (β_1) will be attenuated to the extent that the tasks used are impure and unreliable.

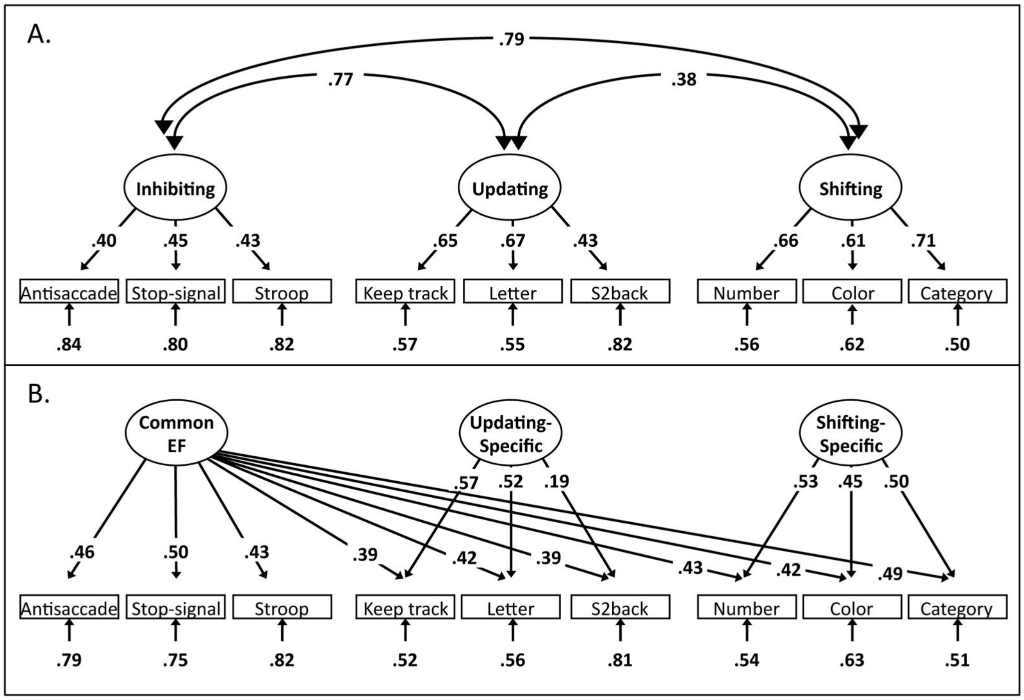

Figure 2.

Correlated factors (panel A) and bifactor (panel B) parameterizations of the unity/diversity model of executive functions (EFs). Ellipses depict EF latent variables that explain correlations among the observed tasks (depicted by rectangles). Numbers on single-headed arrows are standardized factor loadings, numbers at the end of arrows are residual variances (variance not explained by the EFs), and numbers on double-headed arrows are latent variable correlations. In panel A, each EF predicts one task, and the three EFs are allowed to correlate. Their significant correlations indicate that they capture something common (unity), but the fact that these correlations are not perfect suggests that they also capture different processes (diversity). In panel B, this unity and diversity is captured by latent variables rather than by correlations: the Common EF factor predicts all nine tasks, capturing unity, and the Updating- and Shifting-Specific factors capture additional variance in the updating and shifting tasks, once the common factor is removed. Parameters taken from Friedman, Miyake, Robinson, and Hewitt (2011); all $p<.05$.