Psychology not Educational Neuroscience is the way forward for improving educational outcomes for all children: Response to Gabrieli (2016) and Howard-Jones et al. (2016)

Jeffrey S. Bowers

University of Bristol

Author Note:

Jeffrey S. Bowers, School of Experimental Psychology, University of Bristol.

Acknowledgments: I would like to think Patricia Bowers and Jacqueline Aldridge for comments on previous drafts.

Correspondence concerning this article should be addressed to Jeffrey S Bowers, School of Experimental Psychology, 12a Priory Road, Bristol, BS8-1TU. Email j.bowers@bristol.ac.uk

Abstract

In Bowers (2016) I argued that there are practical problems with educational neuroscience (EN) that explain why there are no examples of EN improving teaching, and principled problems with the logic motivating EN that explain why there likely never will be. In the following I consider the main responses raised by both Gabrieli (this issue) as well as Howard-Jones et al. (this issue) and find them all unconvincing. Following this exchange there are still no examples of EN providing new insights to teaching in the classroom, there are still no examples of EN providing new insights to remedial instructions for individuals, and as I detail here, there is no evidence that EN is useful for the diagnosis of learning difficulties. The authors have also failed to address the reasons why EN is unlikely to benefit educational outcomes in the future. Psychology, by contrast, can (and has) made important discoveries that can (and should) be used to improve teaching and diagnostic tests for learning difficulties. This is not a debate about whether science is relevant to education, but what sort of science is relevant.

Keywords: educational neuroscience; education; instruction; neuroscience; mind, brain, and education.

Psychology not Educational Neuroscience is the way forward for improving educational outcomes for all children: Response to Gabrieli (2016) and Howard-Jones et al. (2016)

In Bowers (2016) I argued that there are practical problems with educational neuroscience (EN) that explain why there are no examples of EN improving teaching, and principled problems with the logic motivating EN that explain why there likely never will be. Gabrieli (this issue) and Howard-Jones et al. (this issue) raise a number of objections to my characterization of EN, including the fact that the goals of EN are more general than improving teaching. I will respond to these points, but the key take-home points I would emphasize from the start are that the authors did not provide any good examples of EN improving classroom or remedial instruction, and they did not provide any reasons to believe that EN research will span the “bridge too far” (Bruer, 1997) in the future.

Before responding to the two commentaries in turn, I want to emphasize that my critique of EN should not be taken as an argument against neuroscience, nor as a claim that education should be uninformed by science. Rather, my main claim is that neuroscience (or cognitive neuroscience) is unlikely to be useful in designing or evaluating teaching methods for the classroom, or for interventions for individuals with learning difficulties. Furthermore, as I emphasize here, there is little reason to think that neuroscience will improve educational outcomes through the early diagnosis of learning disorders. The core problem is that for these purposes neuroscience does not add anything above and beyond psychology. Psychology, by contrast, can (and has) made important discoveries that can (and should) be used to improve teaching and diagnostic tests for learning difficulties. This is not a debate about whether science is relevant to education, but what sort of science is relevant.

Gabrieli (2016):

Gabrieli agrees with me that EN has yet to make any practical benefits for teaching in the classroom, but he does not see this as a serious problem given that he views EN as basic research enterprise that addresses topics of importance to education. He sees EN as analogous to cognitive neuroscience, or affective neuroscience, or social neuroscience that characterize the functional brain organization of perception, learning and memory, thought, emotion, and social cognition. And just as advances in cognitive, affective, and social neuroscience are expected to contribute to treatments for a range of disorders, including schizophrenia, depression, anxiety, and autism, advances in EN are expected to improve teaching in the future.

I see two problems with this characterization of EN. First, in contrast with Gabrieli, many researchers claim that EN has already helped improve teaching in the classroom, and the explosion of articles, journals, research centers, teaching programs, and funding directed at EN reflect the widespread view that EN is poised to further advance teaching in the near-term. I gave a number of quotes to this effect in the target article, and indeed, Howard-Jones et al. summarize multiple examples of EN taken to be relevant to informing instruction today (see below). Indeed, the push to train teachers about neuroscience (see target article) only makes sense on the view that EN can improve teaching in the classroom now.

Second, and more importantly, Gabrieli's optimism regarding the promise of improving teaching through EN is based on a flawed an analogy to cognitive, social, and affective neurosciences. It does indeed seem plausible that these areas of neuroscience may in the future improve the wellbeing of persons with learning difficulties, depression, anxiety disorders, schizophrenia, etc., and indeed, neuropharmacology has already lead to important clinical treatments for learning difficulties, depression, anxiety disorders, schizophrenia, etc. The critical point to note, however, is that in all cases, the clinical advances were the direct result of brain intervention (e.g., Ritalin for attention deficit hyperactivity disorder, antidepressants, antianxiety, antischizophrenic drugs). This is qualitatively different to the claim that an improved understanding of the brain will inform behavioral interventions (Bishop, 2013, makes the same point). Different levels of analyses support different levels of intervention: A better understanding of genetics will likely support new genetic interventions (some relevant to education), a better understanding of the brain will likely support new medical interventions (some relevant to education), and a better understanding of psychology will likely support new behavioral interventions (some relevant to education). And of course, linking levels of description is a fundamental goal of science (e.g., understanding the neurobiological and genetic basis of dyslexia is important). But whether neuroscience (or cognitive neuroscience) can contribute above and beyond psychology to designing and evaluating behavioral intervention is anything but clear. The analogy to medical interventions is common (also see initial quote by Howard-Jones et al.) but misguided.

Gabrieli and I also disagree about the current success and future promise of EN for individualized instruction and prediction. According to Gabrieli, a key achievement of EN is that it improves our ability to predict whether a child should be given a specific type of instruction or is likely to have difficulties in the future. Gabrieli describes a number of studies that are taken as strong evidence for the success of neuroscience in making predictions: for example, diagnosing reading disorders or predicting whether a specific math tutoring intervention is well suited for specific children.

Indeed, there is now a large literature and long history of using brain imaging to diagnose future learning difficulties and longitudinal outcomes across a range of domain, including reading (e.g., Hoeft et al., 2007), language more generally (Guttorm et al., 2005), second language learning (Qi et al., 2014), mathematics (Dumontheil & Klingberg, 2012), future working memory capacity (Ullman et al., 2014), amongst others domains. In an excellent review article, Gabrieli et al. (2015) reviewed 72 neuroimaging studies used to make predictions for a wide range of behavioral outcomes. A key finding from this review is that 61 of the 72 identified studies were correlational in that brain measures were associated with known behavioral outcomes. As noted by Gabrieli et al. (2015), the success of these studies to predict the outcomes is better described as "post-diction" – akin to correctly predicting stock market crashes in the past rather than in the future. This greatly inflates the estimates of prediction, especially when post-diction is based on studies with small sample sizes, that include a wide range possible variables (brain measures) for prediction, and where the authors have a great deal of analytical flexibility (as typical in these studies).

One way to provide a better estimate of prediction is "cross validation" in which a subset of data from a given study are used to develop a model that can be used for prediction, and then this model is used with the remaining data in order to make predictions. Gabrieli et al. (2015) identify four such studies relevant to education that employed some form of cross validation, three concerned with reading, one with math, and only one of these studies was carried out on young children that could in principle be used in to inform early diagnosis (Bach et al., 2013). However, the authors themselves urged caution in interpreting these results given the extremely small sample size in the study (11 normal and 6 poor readers). Indeed, underpowered studies that report significant effects are the least likely to replicate (e.g., Button et al., 2013).

As noted by Gabrieli et al. (2015), cross-validation is no substitute for making predictions on an independent sample. Indeed, these predictions are only useful to the extent that someone else can successfully predict learning outcomes in another lab with a comparable sample of participants. But despite over two decades of using neuroscience to make predictions, not a single study relevant to education has been carried out (and only 1 out of 72 studies in the Gabrieli et al. review). Before making any strong claims regarding the predictive successes of EN it is important that replications on independent samples are carried out.

Even if EN is shown to support predictions in independent samples of children, two more hurdles must be passed before the results are useful for education. First, the predictions from imaging are only relevant to educators if they contribute something above and beyond the predictions based on behavioral measures. However, in most (or all) cases in which imaging studies are claimed to improve prediction above behavioral measures alone, some relevant behavioral measures were not included. Consider reading, the domain in which the most prediction studies have been carried out. In a recent behavioral study, Heath et al. (2014) assessed the pre-literacy skills of 102 four-year olds at the beginning of Preschool and then assessed rapid automatized naming at the beginning and end of kindergarten. Children's literacy outcomes were measured at the end of Year 1 (age 6-7). Combining these behavioral measures with a range of family factors, including parents' phonological awareness, provided accurate classification of reading difficulties (sensitivity = .85; specificity = .90; overall correct = .88). Nevertheless, few neuroimaging studies concerned with the prediction of reading have considered whether imaging outperforms behavioral studies when family factors are considered, and no study has considered parents' phonological awareness, a powerful predictor in this study. This again suggests that the unique contribution of neuroscience to prediction has been overestimated.

Second, even if we assume that the predictions (or post-dictions) from neuroimaging studies are not over-estimated, and further, we assume that neuroimaging tools will continue to provide significant predictions above and beyond the best behavioral measures, there are good reasons to question the value of making earlier diagnoses. Again consider the domain of reading where is it frequently claimed that early diagnosis is key. A question that needs to be asked is how early? Behavioral measures appear to do a good job in making predictions in kindergarten (Heath et al., 2014). Would it be helpful to diagnosis likely dyslexia earlier, before children are introduced to written letters and words? This seems to be the assumption of various EN studies that have attempted to diagnosis dyslexia as early as 36 hours after birth (e.g., Molfese, 2000). But given that the best forms of interventions involve training skills directly relevant to literacy (e.g., P. Bowers, Kirby, & Deacon, 2010; Snowling & Hulme, 2012), it is not clear what benefits can be gained from interventions carried out in younger children. Indeed, there is little evidence that interventions focusing on non-written materials are successful. For example, the intervention FastForword is based on the claim that dyslexia is the result of rapid auditory temporal processing skills that can be trained before children are introduced to written letters and words. However, the main evidence for the efficacy of FastForword comes from flawed neuroimaging studies (Bishop, 2013). When a systematic review of behavioral studies was carried out there is no evidence that FastForword improves reading (Strong, Torgerson, Torgerson, & Hulme, 2011).

In sum, Gabrieli provides no good reasons to support the claim that EN will be useful for classroom interventions in the future, no evidence that EN has improved the predictions of learning disorders above and beyond behavioral data given that there are no examples of neuroscience predicting learning disorders in independent samples of children, and no reasons to think that earlier predictions would improve educational outcomes in any case.

Howard-Jones et al.:

Howard-Jones et al. are more critical of my critique of EN, and more positive about the current successes of EN in informing teaching in the classroom. Indeed, they identify a number of EN studies that they claim have already provided important insights for teaching. However, their commentary reveals a number of theoretical confusions regarding my position, and their examples of EN contributing useful insights for instruction do not bear up to scrutiny.

Theoretical confusions

According to Howard-Jones et al., I am attacking a straw man when I challenge the view that neuroscience alone is useful for education. They point out that the “neuroscience” in EN refers to cognitive neuroscience that is concerned with making links between the brain and behavior. As a consequence, psychology is central to EN. My focus on the neuroscience to the exclusion of other fields is thought to undermine my critique of EN.

But this reveals a fundamental confusion regarding my position. My claim is that the neuroscience in cognitive neuroscience does not contribute anything to education above and beyond psychology. As I wrote:

> Even areas of neuroscience that focus on both brain and behavior, such as cognitive neuroscience, are only relevant to education to the extent that they provide new insights into behavior. (no page number yet)

This conclusion was supported by my detailed review of the literature where I failed to identify any examples of the neuroscience in EN contributing any novel insights into the design or evaluation of teaching in the classroom, or any insights for interventions for individuals with learning disorders. Below I consider the examples of EN that Howard-Jones highlight and show why these studies also fail to address my critique.

Another (related) confusion by Howard-Jones et al. is expressed in their passage:

> In Bowers (2016), psychology and neuroscience are pitted as competitors in explaining behavior and, using arguments rehearsed by others (Bishop, 2014; Coltheart & McArthur, 2012; Davis, 2004; Schumacher, 2007), it is proposed that psychology should have central status. (no page number yet)

In fact, I agree with Howard-Jones et al. that behavior and brain play essential roles in cognitive neuroscience given that the goal of cognitive neuroscience is to explain the brain basis of behavior. There is no “knowledge hierarchy” in this domain. But again, my claim is neuroscience is irrelevant to the design and evaluation of teaching. This is why I suggested replacing the “bridge too far” analogy with the “one-way street” analogy: The neuroscience in EN is of interest to neuroscientists and psychologists, and it may contribute to future medical advances relevant to education, but it is (and likely always will be) irrelevant to the task of designing or evaluating instruction.

A third theoretical confusion used to argue for the promise of EN is expressed in the following passage:

> …if one accepts that psychological theory can contribute to educational practice and that neuroscience can contribute to psychological theory, it is illogical to disallow the transitive inference and to instead argue that, in principle, neuroscience is irrelevant to educational practice.

But there is nothing transitive about this argument: Neuroscience might improve psychological theories in ways that are irrelevant to education. Indeed, given the different levels of analysis involved in brain and behavioral research (as noted in my response to Gabrieli), there is no reason to assume that the neuroscience in EN will add any new insights relevant to teaching. What would be relevant are examples of neuroscience providing unique insights relevant to designing or evaluating instruction in the classroom or for individuals, or to early diagnosis of learning disorders, or at minimum, good reasons to think there is promise of achieving these goals in the future. Thus it is interesting that Howard-Jones et al. have provided specific examples of EN research contributing insights relevant to education. However, I would characterize these studies in the same way as the studies I reviewed in the target article; namely, they are either trivial in the sense that the finding are self-evident, misleading in the sense that behavioral studies already motivated the interventions, or unwarranted in the sense that the recommendations do no follow from the EN study.

A review of the evidence cited by Howard-Jones et al.

Under the heading “Examples illustrating the EN approach and its potential” Howard-Jones first cite Tanaka et al. (2011), as did Gabrieli in his commentary. This study provides neuroscientific evidence against the IQ-discrepancy criteria for diagnosing dyslexia given that discrepant and non-discrepant poor readers show similar brain abnormalities. As noted by Gabreili, this finding has important policy implications, as it suggests that poor readers ought to receive remedial support regardless of IQ. I agree with this, but it is important to note that behavioral arguments against the IQ-discrepancy definition of dyslexia have a long history (e.g., Stanovitch, 1991), and Tanaka et al. motivate their neuroimaging study by a long list of behavioral studies that reported similar phonological deficits in both discrepant and non-discrepant poor readers, and critically, cite multiple behavioral studies have shown that both types of readers respond similarly to remedial reading programs designed to ameliorate phonological deficits. This later finding is far more relevant to policy than the (interesting) finding that discrepant and non-discrepant poor readers show similar brain anomalies. Indeed, like many of the studies I review below, this is a good example of the “one-way street” in which EN is relevant to neuroscientists and psychologists but provides no new suggestions for improving interventions that would be relevant to teachers.

Next, Howard-Jones et al. cite a successful intervention study for dyslexia by Gonzalez et al. (2015) that was motivated by neuroscience research reported by Froyen et al. (2008). Froyen et al. reported electrophysiological research showing abnormal processing of letter-sound correspondences in dyslexics, and critically, Howard-Jones et al claim that this study showed that the deficit persists longer than was previously suggested by behavioral data. However, Foyen et al. (2008) cited behavioral evidence suggesting just this, and the intervention used by Gonzalez et al (2015) study was a close adaptation of a previous study motivated by behavioral data (Tijms, 2007).

Next, Howard-Jones et al. cite an intervention by Tsang, Blair, Bofferding, and Schwarz (2015) designed to teach children negative number concepts that was motivated by various neuroscience results. But as Howard-Jones et al. note themselves, the initial hypothesis was motivated by behavioral studies. The same is true for the Number Race and Graphogame interventions that are designed to train approximate comparison processes on the one hand, and learning exact numerosities and symbol knowledge on the other. Once again, behavioral evidence relevant to designing these instructions preceded the neuroscience (e.g., see Wilson, Revkin, Cohen, Cohen, & Dehaene, 2006). For early behavioural evidence in support of the claim that instruction should involve comparing numerical magnitudes see Griffin, Case, and Siegler (1994).

It is not a coincidence that all these EN studies were carried out after the relevant hypothesis was advanced on the basis of behavioral data. Indeed, the goal of cognitive neuroscience is to understand how the brain supports psychological processes, and accordingly, psychology comes first. As Howard-Jones et al. write themselves: "…when Bowers claims that instruction investigated by EN is often first motivated by behavioral data, we sincerely hope this is true and continues to be so". This does raise the issue of how EN is going to help motivate new forms of instruction above and beyond psychology.

One possible response to the observation that psychology comes first is that the neuroscience provides stronger evidence for a given hypothesis. On this view, EN provides a stronger motivation to design an intervention. This seems to be the implication of the following point:

> Bowers is completely uncritical of behavioral evidence. For example, he does not consider that many measures of behavior are often unreliable and lack validity (e.g. Holloway & Ansari, 2009; Maloney, Risko, Preston, Ansari, & Fugelsang, 2010; Stolz, Besner, & Carr, 2005) or that the absence of a behavioral change on a psychological measure of behavior does not imply that no change in behavior has occurred. By arguing that neuroimaging has nothing to add if behavior does not change, Bowers makes a classic misinterpretation of null findings.

I agree that there is a replication crisis in psychology, but the problem is worse in cognitive neuroscience. The root causes of the replication crisis are the same in both domains: studies with small sample sizes, single experiments without replication, analytical flexibility (that can lead to p-hacking), and a bias to publish significant results. In contrast to psychological studies, neuroscience studies tend to have smaller sample sizes, fewer replications, and much more analytical flexibility. The consequence is a replication crisis in neuroscience as well (Button et al., 2013). Indeed, this helps explain why many EN results claimed to support specific forms of instruction are not supported when assessed with more powerful behavioral experiments. For example, Bishop (2013) reviewed all the studies between 2003-2011 that measured brain function before and after an intervention that focused on improving language skills. All six studies were seriously flawed (e.g., underpowered, analytical flexibility, no replication) and the conclusions were not supported in multiple and more powerful behavioral studies.

And more importantly, even if the neuroscience results are robust and reliable, the link between the neuroscience and the cognitive processes that support a given skill is often indirect and hard to specify. For example, abnormal BOLD signals in visual cortex may reflect abnormal processes in another brain area (Dorjee & Bowers, 2012), and indeed, as noted in the target article, Olulade Napoliello, and Eden (2013) attributed changes in the BOLD signal in V5/MT (visual cortex) to improvements in phonological processing. If the goal is to identify the cognitive processes that are responsible for learning deficits, then it is almost always more straightforward to test behavior. For instance, in order to test the hypothesis that phonological processes are impaired in dyslexia, rather than looking at BOLD signals in V5/MT (involved in visual motion processing), or temporoparietal cortex (involved in phonological processes), or ventral occipito-temporal cortex (the visual word form area), researchers should measure the phonological skills of children who are struggling to read.

Howard-Jones et al.'s additional claim that I have made a "classic misinterpretation" of null findings reflects another fundamental misunderstanding of my position. If a given behavioral study fails to improve performance on some task then I agree it does not rule out the possibility that the intervention was effective. The study may have been underpowered, or the wrong behavioral test may have been administered. However, these type-1 errors cannot be rectified with neuroimaging. Putting aside the dangers of false positives in neuroimaging results (given the reasons noted above), an intervention is only useful to the extent that it produces functionally relevant behavioral outcomes. If a researcher is concerned that a behavioral intervention was successful but nevertheless null results were obtained, the only appropriate response is to carry out a better and more powerful behavioral study. As I wrote in the target article, brain measures should not even be used as a more sensitive measure of behavioral change:

> Rather than using (expensive) fMRI to look for a weak training effect that may or may not manifest itself in behavior following further training, researchers should explore whether more extensive (cheaper) behavioral training does indeed result in useful changes in behavior. (no page number)

Continuing on with examples of the successes of EN, Howard-Jones et al. describe how EN can provide evidence for the compensatory processes in remediation, with implications for designing instruction. The authors cite a study that found regions in the right prefrontal cortex to be more activated after a phonological remediation for dyslexia (Shaywitz et al., 2004). Based on this observation, Howard-Jones et al. argue that a better understanding of these brain regions will contribute to future interventions.

But the strategy of studying the prefrontal cortex in an effort to design better forms of interventions is misguided given that these regions are undoubtedly involved in supporting many different functions (any one of which may have contributed to these results), characterizing these functions is no easy matter, and it is not clear that characterizing these functions will provide any new insights into instruction. To make things worse, as summarized in the target article, the pattern of BOLD changes in this study were more complicated than summarized by Howard-Jones et al., making it all the more unclear what teaching implications follow from this study.

Psychological research, by contrast, leads to highly specific hypotheses regarding possible compensatory strategies that may be useful for dyslexics. For example, we know that learning is best when information is studied in a meaningful and organized way (Bower, Clark, Lesgold, and Winzenz, 1969), and we also know that the writing system is English is highly organized and logical when morphological and etymological constraints are considered in addition to phonology (Venezky, 1967). This suggests that we might design literacy interventions that highlight the meaning and organization of English spellings rather than focus so heavily on meaningless letter-sound correspondences (phonics); what P. Bowers and Kirby (2010) called “structured word inquiry”. Consistent with this hypothesis, there is now growing evidence that teaching children the logic of their writing system helps reading and writing (e.g., Devonshire, Morris, Fluck, 2013), and vocabulary (e.g., P. Bowers & Kirby, 2010), especially in younger and struggling children. It is hard to imagine how cognitive neuroscience could have led to this hypothesis. Of course, more experimental work is needed before any strong conclusions are warranted regarding structured word inquiry, and this will require randomized control trials that measure reading skills.

Advocates of EN often do not appreciate the challenge that compensatory processes pose for EN. For example, consider the observation that dyscalculia is often associated with abnormal activation of the intrparietal cortex. This finding is taken as evidence that dyscalculia is associated with in impaired “number sense”. This in turn is thought to have implications for education. For example, Butterworth et al. (2011) conclude “[a]lthough the neuroscience may suggest what should be taught, it does not specify how it should be taught” (p. 1051). Howard-Jones et al. again highlight the importance of this work because it identifies <u>what</u> should be taught (the authors suggest that other forms of research can build on EN studies to identify <u>how</u>

to best achieve this). However, this conclusion rests on the standard logic of EN; namely, the claim that interventions should target the deficit. An equally plausible hypothesis is that interventions should target compensatory skills. That is, the neuroscience does not indicate what should be taught. Behavioral studies are needed to distinguish between these two possibilities. And in any case, the hypothesis that dyscalculia is associated with a deficit in the "number sense" was first supported on the basis of behavioral studies (e.g., Koontz & Berch, 1996; Landerl, Bevan, Butterworth, 2004).

Finally, Howard-Jones et al. describe evidence for the importance of teaching children about the plasticity of the brain. More specifically, the authors claim that a student's theory of learning can be influenced by their views of the brain, and further, brain theories of learning can be an important determinant of academic motivation and success. The first claim is based on a study that found that children judged that the intelligence is more malleable in a multiple-choice test after being taught that the brain is plastic (29% vs. 21% of children endorsed the statement "you can always change how intelligent you are" after the lesson; Dekker & Jolles, 2015). This is not a surprising finding (and furthermore, it is not clear what the right answer is). The later two claims are based on studies that found teaching brain plasticity to students improved their academic motivation as measured by teachers (Blackwell, Trzesniewski, & Dweck, 2007), and raised the grades of students most likely to drop out of school (Paunesku et al., 2015). These later findings are striking, and replications would be warranted. But in any case, these studies provide no evidence for EN. In order to support the conclusion that learning about the brain is relevant to these findings it is necessary to show that the experimental group shows the same advantage compared to a control group who were taught about the psychology of learning and evidence that adults can learn (without reference to the brain). This was not done.

It is perhaps worth noting that I am not claiming that it is logically impossible for EN to improve teaching in the future. Rather, I am claiming that the principled problems with the logic motivating EN makes it unlikely. It is always possible that a neuroscience finding will lead to a novel hypothesis regarding how cognitive processes support a skill (or how cognitive processes are impaired in the case of a deficit), and this hypothesis in turn suggests a new form of instruction, and this instruction is demonstrated to be useful in subsequent behavioral work (a case where the neuroscience precedes the behavior studies). But there are no examples of this thus far, and it is hard to even imagine such an outcome. Consider Gabrieli's first example of neuroimaging constraining theory in a relevant manner for education; namely studies suggesting that dyslexia reflects a deficit in accessing phonetic representations rather than a deficit in the quality of the representations themselves (e.g., Boets et al., 2013). For the sake of argument, put aside the fact that this hypothesis was first advanced on the basis of behavioral work (Ramus & Szenkovits, 2008). The question still remains, what does an educator do with this hypothesis? It is not at all obvious how an access deficit motivates a new form of instruction, let alone whether instruction should be designed to ameliorate or compensate for this deficit. This is just one example, but the failure of EN to generate any novel hypotheses that suggest new forms of instruction is telling. Far more promising is to develop and test hypotheses on the basis of behavioral work, and indeed, here, there are already many examples of psychological research doing just this (e.g., Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013; Roediger, Finn, & Weinstein, 2012).

Setting the record straight

Let me briefly respond to the claim that I mischaracterized Goswami (2004a) in one of multiple "unfortunate use of quotes". The quote in question suggests that neuroscience can provide a more direct way of evaluating the efficacy of an intervention than behavior itself, but Howard-Jones et al. claim that the quote was taken out of context. Although I maintain that my reading of the full text is a reasonable interpretation of what was written, I am sorry for the confusion. But in any case, the important point is not whether neuroscience provides a more direct measure of the efficacy of a treatment, but whether neuroimaging should be used at all in evaluating interventions. Goswami (2004a) does appear to endorse that view when she writes: "If the effects generalize to literacy (for example, via improved automaticity), then changes in occipital, temporal and parietal areas should also be observed" (no page yet). And again this view seems to be endorsed by Goswami (2004b) in the quote provided by Howard-Jones et al.: "If an excercise-based package actually improves reading in dyslexic children, there should be measurable effects in the neural systems for reading" (p 179). These statements are simply false. For example, another possible outcome is that reading improves following exercise and it reflects compensatory processes that occur in brain regions not normally associated with reading, such as in the right prefrontal cortex. Changes in the neural systems typically used for reading cannot be used to assess whether the treatment worked.

Similarly, although Howard Jones et al. write that "EN does not claim… that educational achievement should be evaluated using brain imaging", they appear to claim that that neuroscience is relevant to evaluating interventions in other contexts. For example, as detailed above, the authors state that EN is useful when behavior measures fail to observe an effect following an intervention. The implication is that brain imaging can be used to measure behavioral change that was not captured in task performance. More importantly, as noted in the target article, the claim that neuroscience is relevant to evaluating an intervention is common in the literature, but it is incorrect (see Coltheart & McArthur, 2012).

Summary

In Bowers (2016) I claimed that EN has not contributed any novel insights into instruction above and beyond the behavioral research, and there are still no examples after this exchange. Similarly, as detailed here, there are no good examples of the neuroscience in EN uniquely contributing to the early diagnosis of learning disorders or predicting longitudinal outcomes of a given intervention for specific children. Perhaps most importantly, neither Gabrieli nor Howard-Jones et al. have provided any sound reasons to think that the neuroscience in EN will contribute to improving instruction in the future. Given all the excitement and funding directed at EN, some educators and researchers might find these conclusions surprising.

I want to emphasize again that I am not arguing that science is irrelevant to improving education, nor am I claiming that understanding cognition is irrelevant (as Howard-Jones et al. suggest). Rather, I am claiming that psychology is the relevant discipline to improve educational outcomes for all children.

References

Bach, S., Richardson, U., Brandeis, D., Martin, E., & Brem, S. (2013). Print-specific multimodal brain activation in kindergarten improves prediction of reading skills in second grade. *Neuroimage*, *82*, 605-615.

Bishop, D. V. (2013). Research review: Emanuel Miller Memorial Lecture 2012—Neuroscientific studies of intervention for language impairment in children: Interpretive and methodological problems. *Journal of Child Psychology and Psychiatry*, *54*, 247–259. http://dx.doi.org/10.1111/jcpp.12034

Bishop, D. V. M. (2014). What is educational neuroscience? Retrieved from *https://figshare.com/articles/What_is_educational_neuroscience_/1030405*

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievment across an adolescent transition: A longtitudinal study and an intervention. *Child Development*, *78*, 246-263.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, *8*, 323–343. http://dx.doi.org/10.1016/S0022-5371(69)80124-6

Bowers, J.S. (2016). The practical and principled problems with educational neuroscience. *Psychological Review*. http://dx.doi.org/10.1037/rev0000025

Bowers, P. N., & Kirby, J. R. (2010). Effects of morphological instruction on vocabulary acquisition. *Reading and Writing: An Interdisciplinary Journal*, *23*, 515–537.

Bowers, P. N., Kirby, J. R., & Deacon, S. H. (2010). The Effects of Morphological Instruction on Literacy Skills: A Systematic Review of the Literature. *Review of Educational Research*, *80*, 144–179.

Bruer, J. T. (1997). Education and the brain: A bridge too far. *Educational Researcher, 26,* 4–16.

Butterworth, B., Varma, S., & Laurillard, D. (2011). Dyscalculia: From Brain to Education. *Science, 332(6033*), 1049-1053. doi:10.1126/science.1201536

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*,*14*(5), 365-376.

Coltheart, M., & McArthur, G. (2012). Neuroscience, education and educational efficacy research In S. Della Sala & M. Anderson (Eds.), *Neuroscience in Education: The good, the bad and the ugly* (pp. 215-221). Oxford: Oxford University Press

Davis, A. J. (2004). The Credentials of Brain-based Learning. *Journal of Philosophy of Education, 38*, 21-36.

Dekker, S., & Jolles, J. (2015). Teaching About "Brain and Learning" in High School Biology Classes: Effects on Teachers' Knowledge and Students' Theory of Intelligence. *Frontiers in Psychology, 6, 8*. doi:10.3389/fpsyg.2075.01848

Devonshire, V., Morris, P., Fluck, M. (2013). Spelling and reading development: the effect of teaching children multiple levels of representation in their orthography. *Learning and Instruction, 25*, 85-94.

Dorjee, D., & Bowers, J. S. (2012). What can fMRI tell us about the locus of learning? *Cortex, 48*, 509–514. http://dx.doi.org/10.1016/j.cortex.2011.06.016

Duff, F. J., Nation, K., Plunkett, K., & Bishop, D. V. (2015). Early prediction of language and literacy problems: is 18 months too early?. *PeerJ*, *3*, e1098.

Dumontheil, I., & Klingberg, T. (2012). Brain activity during a visuospatial working memory task predicts arithmetical performance 2 years later. *Cerebral Cortex*, *22*(5), 1078-1085.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning tech- niques: Promising directions from cognitive and educational psychol- ogy. *Psychological Science in the Public Interest, 14,* 4–58. http://dx .doi.org/10.1177/1529100612453266

Froyen, D. J. W., Bonte, M. L., van Atteveldt, N., & Blomert, L. (2008). The Long Road to Automation: Neurocognitive Development of Letter–Speech Sound Processing. *Journal of Cognitive Neuroscience, 21*, 567-580. doi:10.1162/jocn.2009.21061

Gabrieli, J.D. (2016). The promise of educational neuroscience: Comment on "The practical and principled problems with educational neuroscience". *Psychological Review*.

Gabrieli, J. D., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a humanitarian and pragmatic contribution from human cognitive neuroscience. *Neuron*, *85*, 11-26.

Gonzalez, G. F., Zaric, G., Tijms, J., Bonte, M., Blomert, L., & van der Molen, M. W. (2015). A Randomized Controlled Trial on The Beneficial Effects of Training Letter-Speech Sound Integration on Reading Fluency in Children with Dyslexia. *PLoS ONE, 10(12*). doi:10.1371/journal.pone.0143914

Goswami, U. (2004a). Neuroscience and education. *British Journal of Educational Psychology, 74*, 1-14.

Goswami, U. (2004b). Neuroscience, education and special education. *British Journal of Special Education, 31*, 175 - 183 doi:10.1111/j.0952-3383.2004.00352.x

Griffin, S., Case, R., & Siegler, R. (1994). Rightstart: Providing the central conceptual prerequisites for first formal learning of arithmetic to students at-risk for school failure. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice* (pp. 24-49). Cambridge, MA: Bradford Books MIT Press.

Guttorm, T. K., Leppänen, P. H., Poikkeus, A. M., Eklund, K. M., Lyytinen, P., & Lyytinen, H. (2005). Brain event-related potentials (ERPs) measured at birth predict later language development in children with and without familial risk for dyslexia. *Cortex*, *41*, 291-303.

Heath, S. M., Bishop, D. V., Bloor, K. E., Boyle, G. L., Fletcher, J., Hogben, J. H., ... & Yeong, S. H. (2014). A spotlight on preschool: The influence of family factors on children's early literacy skills. *PloS one*, *9*, e95255.

Hoeft, F., McCandliss, B. D., Black, J. M., Gantman, A., Zakerani, N., Hulme, C., ... & Gabrieli, J. D. (2011). Neural systems predicting long-term outcome in dyslexia. *Proceedings of the National Academy of Sciences*, *108*, 361-366.

Hoeft, F., Meyler, A., Hernandez, A., Juel, C., Taylor-Hill, H., Martindale, J. L., ... & Deutsch, G. K. (2007). Functional and morphometric brain dissociation between dyslexia and reading ability. *Proceedings of the National Academy of Sciences*, *104*, 4234-4239.

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children's mathematics achievement. *Journal of Experimental Child Psychology, 103*, 17-29. doi:10.1016/j.jecp.2008.04.001

Howard-Jones, P., Varma, S. Ansari, D., Butterworth, B. De Smedt, B., Goswami, U., Laurillard, D., & Thomas, M. (in press). The principals and practices of educational neuroscience: Commentary on Bowers (2016). *Psychological Review*.

Koontz, K. L., & Berch, D. B. (1996). Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. *Mathematical Cognition*, *2*, 1–23.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8–9-year-old students. *Cognition*,*93*(2), 99-125.

Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., & Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica*, *134*, 154-161. doi:http://dx.doi.org/10.1016/j.actpsy.2010.01.006

Molfese, D. L. (2000). Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain and language*, *72*(3), 238-245.

Olulade, O. A., Napoliello, E. M., & Eden, G. F. (2013). Abnormal visual motion processing is not a cause of dyslexia. *Neuron*, *79*, 180–190. http://dx.doi.org/10.1016/j.neuron.2013.05.002

Paunesku, D., Walton, G. M., Romero, C., Smith, E. N., Yeager, D. S., & Dweck, C. S. (2015). Mind-Set Interventions Are a Scalable Treatment for Academic Underachievement. *Psychological Science*, *26*, 784-793. doi:10.1177/0956797615571017

Qi, Z., Han, M., Garel, K., San Chen, E., & Gabrieli, J. D. (2015). White-matter structure in the right hemisphere predicts Mandarin Chinese learning success. *Journal of Neurolinguistics*, *33*, 14-28.

Roediger, H. L., Finn, B., & Weinstein, Y. (2012). Applications of cognitive science to education. In S. Della Salla & M. Anderson (Eds.), *Neuroscience in education: The good, the bad and the ugly* (pp. 128– 151). New York, NY: Oxford University Press. http://dx.doi.org/10 .1093/acprof:oso/9780199600496.003.0037

Schumacher, R. (2007). The brain is not enough: Potential and limits in integrating neuroscience and pedagogy. Analyse & Kritik, 29(1).

Snowling, M. J., & Hulme, C. (2012). Interventions for children's language and literacy difficulties. *International Journal of Language & Communication Disorders*, *47*, 27–34.

Stanovich, K. E. (1991). Discrepancy definitions of reading disability: Has intelligence led us astray?. *Reading Research Quarterly*, 7-29.

Stolz, J. A., Besner, D., & Carr, T. H. (2005). Implications of measures of reliability for theories of priming: Activity in semantic memory is inherently noisy and uncoordinated. *Visual Cognition, 12*, 284- 336. doi:10.1080/13506280444000030

Strong, G. K., Torgerson, C. J., Torgerson, D., & Hulme, C. (2011). A systematic meta- analytic review of evidence for the effectiveness of the ‘Fast ForWord’language intervention program. *Journal of Child Psychology and Psychiatry*, *52*, 224-235.

Tanaka, H., Black, J. M., Hulme, C., Stanley, L. M., Kesler, S. R., Whitfield-Gabrieli, S., . . . Hoeft, F. (2011). The Brain Basis of the Phonological Deficit in Dyslexia

Is Independent of IQ. *Psychological Science, 22*, 1442-1451. doi:10.1177/0956797611419521

Tijms, J. (2007). The development of reading accuracy and reading rate during treatment of dyslexia. *Educational Psychology*, *27*, 273-294.

Tsang, J. M., Blair, K. P., Bofferding, L., & Schwartz, D. L. (2015). Learning to "See" Less Than Nothing: Putting Perceptual Skills to Work for Learning Numerical Structure. *Cognition and Instruction*, *33*, 154-197. doi:10.1080/07370008.2015.1038539

Ullman, H., Almeida, R., & Klingberg, T. (2014). Structural maturation and brain activity predict future working memory capacity during childhood development. *The Journal of Neuroscience*, *34*, 1592-1598.

Venezky, R. (1967). English orthography: Its graphical structure and its relation to sound. *Reading Research Quarterly, 2*, 75–105.

Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., & Dehaene, S. (2006). An open trial assessment of" The Number Race", an adaptive computer game for remediation of dyscalculia. *Behavioral and Brain Functions*, *2*, 1.