IJEP
International Journal of Educational Psychology

Volume 7, Number 3

Hipatia Press
www.hipatiapress.com

Articles

Reviews

Instructions for authors, subscriptions and further details:

http://ijep.hipatiapress.com

Epigenetics: What do Psychologists Need to Know?

Ben Hayes[1]

1) University College London, UK

Date of publication: October 24th, 2018
Edition period: October 2018 – February 2019

To cite this article: Hayes, B. (2018). Epigenetics: What do Psychologists need to know? *International Journal of Educational Psychology*, *7*(3), 230-247. doi: 10.17583/ijep.2018.2915

To link this article: http://dx.doi.org/10.17583/ijep.2018.2915

IJEP – International Journal of Educational Psychology, Vol. 7 No. 3 October 2018 pp. 230-247

Epigenetics: What Do Psychologists Need to Know?

Ben Hayes
University College London

Abstract

This paper reviews the rapidly developing field of epigenetics, providing an accessible explanation of the key ideas and some illustrative examples of work in the field. Although very much a biological discipline the implications of the developing knowledge in this area are very significant for educational psychologists and this paper aims to provide an introduction to what is becoming a very significant shift in how people think about learning and development. Understanding the processes that underlie epigenetic change and the research that the new knowledge is based on will be important for educational psychologists in order to understand this important developing area of thinking about development and learning. Consensus is growing that intergenerational transmission of epigenetic changes are a reliable phenomenon, establishing the principle of the inheritance of acquired characteristics. This contrasts starkly with models of biological determinism and provides a new way of thinking about educational and societal change.

Keywords: epigenetics; genetics; heritability; biological determinism; development.

2018 Hipatia Press
ISSN: 2014-3591
DOI: 10.17583/ijep.2018.2915

Hipatia Press
www.hipatiapress.com

IJEP – International Journal of Educational Psychology, Vol. 7 No. 3 October 2018 pp. 230-247

Epigenética: ¿Qué Necesitan Saber Los Psicólogos?

Ben Hayes
University College London

Resumen

Este artículo revisa el cambiante campo de la epigenética, aportando una explicación asequible tanto a ideas clave como a ejemplos ilustrativos en esta área. Aunque se trata de una disciplina básicamente biológica, las implicaciones del creciente conocimiento de esta área resultan muy significativas para psicólogos educacionales y este artículo pretende proporcionar una introducción a lo que se está convirtiendo en un cambio significativo acerca de lo que la gente piensa sobre aprendizaje y desarrollo. Entender los procesos que sustentan los cambios epigenéticos y la investigación sobre la que se basan, será de gran importancia para los psicólogos educacionales a la hora de entender esta importante área de pensamiento basada en desarrollo y aprendizaje que está en constante progreso. Hay un creciente consenso que la transmisión intergeneracional de los cambios epigenéticos es un fenómeno fiable, que establece el principio de la heredabilidad de características adquiridas. Esto contrasta claramente con los modelos de determinismo biológico y aporta una nueva vía de pensamiento acerca de los cambios educacionales y sociales.

Palabras clave: epigenética; genética; heredabilidad; determinismo biológico; desarrollo

2018 Hipatia Press
ISSN: 2014-3591
DOI: 10.17583/ijep.2018.2915

In education, the perceptions people have about the impact of nature and nurture on learning and development hold great significance. Our understanding of genetic inheritance and its influence on a child can shape how a psychologist or teacher thinks about their work, it can shape school or college policy and it can shape legislation. A common sense understanding of genetic inheritance currently sits largely with the "Darwinian" tradition, although Darwin himself did not identify genetic mechanisms of course. In this way of thinking a genome is passed from one generation to the next. It is subject to random mutations, which may or may not give the individual an advantage. Selection of the genotypes (and phenotypes) that survive and successfully reproduce is guided by this process of random mutations of the genome and any additional advantage that a mutation in the genetic sequence bestows on the host organism.

Under the paradigm of Darwinian evolution education has a fairly limited role when thinking about change across generations. Educating someone has value for that individual during their lifetime and possibly for the society or community that they are part of as a result of what they can contribute. However, any changes a person makes to themselves (or is subjected to) during their lifetime do not get passed on to subsequent generations. A person who struggled with some aspect of life as a result of their genes might expect their children to experience the same struggle, no matter what that person has done in their life to remove barriers or overcome challenges for themselves. Their children would have to overcome them in the same way, from the same starting point. Our thoughts and actions might form part of the culture or knowledge base of the community the next generation are born into, but that new generation start with the same biological blueprint and have to learn (or re-learn) any behavioral adaptations or other advantages that previous generations acquired. The notion that life outcomes can be attributed to fixed biological factors is strong. In the long-term human beings are tied to processes that have a distinct 'biological determinism' (Lewontin, 1976).

Epigenetics is starting to offer a radically different perspective: that changes we make in our lifetime not only affect us but that these changes can be passed onto subsequent generations through genetic processes as well as through culture and learning. This paper will explore this new thinking, explaining the scientific basis for the ideas.

It will start by describing the ‘problem of missing heritability’ that has been discussed in genetic research, and why is it not possible to explain human diversity through traditional genetics alone. It will then explain the fundamentals of epigenetics and how they potentially help solve this problem before exploring some specific examples of research that have developed our understanding of epigenetics over the last decades.

Finally, the paper will explore some possible implications of these ideas for educators and psychologists in particular. It will offer a new template for thinking about inheritance.

The problem of missing heritability

The human genome project came with great hopes for genetic solutions to many of the problems that human beings face. Heralded as a ‘holy grail’ of scientific achievement (Lock & Palsson, 2016, p. 76) it promised an explanation at an individual level for the diversity in human physiology, behaviour and psychology. Differences in our DNA would be identified and pinpointed so any unusual patterns in the DNA, perhaps caused by a mutation, could be mapped and used to help explain differences between us. Twin study research in particular has indicated that there could be a significant genetic component to psychological phenomenon as broad as happiness for example (Blum et al., 2009).

Twin study research has led to the hypothesis that much variation in human experience can be attributed to our genes. Comparing the life outcomes of genetically identical individuals in similar and dissimilar environments provided seemingly irrefutable evidence that 60% or more of variability between human beings could be attributed to genetic similarity or difference (Boomsma, Busjahn, & Peltonen, 2002).

Genomewide Association Studies (GWAS) take a very different approach to looking at genetic inheritance (Craddock, 2013). GWAS studies map millions of DNA patterns across many thousands of participants. One study for example (Chang et al., 2016) looked at the genetic code of over 2000 participants who had diabetes related cataracts and compared them to nearly 3000 controls. The participants came from a Scottish national dataset on diabetes patients. The results found a specific gene that could be identified as having an effect in the development of cataracts in diabetes.

Twin studies have suggested that variation between human phenotypes is, for many characteristics, largely related to genetic influences. The huge power associated with a GWAS study held great promise for being able to pinpoint the specific genes that would explain specific differences that might be seen between humans. However, there was a problem. GWAS studies have identified over 1200 specific DNA patterns or pairs in the DNA sequence that are associated with 165 human diseases or traits (Zuk et al., 2012) but when the results are extrapolated they only account for 20%-30% of the variability between human beings. So on the one hand twin studies would claim that the majority (60% or more) of human variation is attributable to genetic variation, and on the other hand GWAS studies show that when you pinpoint specific DNA variations in population level data you can only explain 20%-30% of the variation. There was no easy way to explain the difference between the two and the gap has been termed 'the missing heredity problem' (Plonka, 2016; Zuk et al., 2012; Slatkin, 2009). In fact, this crisis in genetic research and the failure to identify how many human traits are directly linked to genetic variation is just the tip of the iceberg. Genetic expression is generally subtle, not specific, and the way that the environment can influence which genes are important and which are not is complex, meaning that there are many anomalies in genetic research. Anomalies where what is predicted by genetic mechanisms is not seen in reality.

Epigenetics has been highlighted as the most likely source for an explanation for what might be going on (Slatkin, 2009; Carey, 2011; Spector, 2012; Lock & Palsson, 2016). So, what is epigenetics?

Epigenetics: An introduction

To explain the significance of the shift in thinking that these ideas bring it is worth taking some time to describe the mechanisms involved.

"When scientists talk about epigenetics they are referring to all the cases where the genetic code alone isn't enough to describe what's happening – there must be something else going on as well" (Carey, 2011, p. 6).

There are many descriptions in the literature of the key epigenetic mechanisms that are involved. Genes are used as the blueprint for cells in the body, which in turn form a blueprint for organs and the body itself. The way that the human body (phenotype) develops is determined in no small part by the genetic code. However, the DNA sequence has to be 'read'. The process of reading a DNA code is controlled by Methyl groups and histones. Histones are coils of proteins that the DNA sequence is wrapped around when stored in a cell. The way in which the histones are attached to the DNA affect whether the gene is switched 'on or off'. This determines whether that piece of DNA has an effect or not on the developing person, and if it does have an effect, how it has that effect. So, in this way the same set of DNA in two different people can produce very different phenotypes.

DNA methylation (DNAm) is the process that wraps the DNA sequence with this additional layer of material that determines how the DNA is read. The DNAm process involves methyl groups being attached to the DNA. Methyl groups are molecules that become markers for other silencing proteins to interact with the DNA, methyl groups are not proteins themselves. These then form groups of proteins, including histones, that moderate, repress or silence the DNA sequences in the genes. Neither the presence of methyl groups or histones change the actual sequence of the DNA itself. They change the way it is used in a cell and how it is read by the body.

Williams and Drake (2015) give a summary of how life experiences, and particularly early life experiences, can 'programme' the genetic code through DNAm. Different patterns of methyl groups can change the organic structure of the body, create long term hormone changes, affect the hypothalamic-pituitary-adrenal axis (associated with many developmental conditions) and influence many other key developmental processes.

Taking a different perspective on the same problem authors writing from an evolutionary perspective exploring the archaeological information about genetics (Brooke & Larsen, 2014) have noted that there is an "established consensus that the essential modelling of the genetic code ended sometime in the Paleolithic". They argue that there can be no meaningful genetic explanations for human behaviors as genetic changes are so slow in terms of their effect on a species. They look to epigenetic changes to explain the way in which humans have responded so quickly and successfully to their environment over the last 10,000 years.

Epigenetics therefore provides the solution to the problem of missing heritability. The additional information present alongside the DNA itself helps explain why GWAS studies have not found that specific genes themselves predict very much of the variance in human life. It is the additional information that goes along with the genes that makes the difference. A fundamental point about DNAm is that how this additional information is wrapped around the DNA is often under our influence or our control. Something that the environment we are in, our life choices and experiences can alter.

It should be noted that recent reviews have highlighted the possibility of other processes and mechanisms that might be involved, and that although well described, the methylation mechanism might be one of several possible epigenetic processes (Scorza et. al., 2018).

Intergenerational transmission

A key aspect of epigenetic processes is that evidence is strongly suggesting that the additional information that surrounds the DNA sequence can be inherited by subsequent generations along with the DNA sequence itself. This finding has been trumpeted by proponents of epigenetics, highlighting the paradigm shift it represents. Ammaniti and Gallese (2014) for example describe the discovery of intergenerational transmission of trauma and resilience as revolutionary in the impact it has on how we understand inheritance, human development and emotional wellbeing. A ground-breaking paradigm shift and ‘revolution’ is also the kind of language used by Carey (2011), Plonka (2016) and Lock and Palsson (2016). These writers argue that epigenetic findings are forcing us to fundamentally re-appraise our thinking about genetic inheritance and the impact of environment on our life outcomes. Despite the rhetoric the shift in thinking has been gradual. Writing only a couple of years before Carey, Slatkin (2009) noted that there was still much to learn about the extent to which epigenetic changes could be inherited, and once inherited whether an epigenetic change would last and persist. Even some recent definitions of epigenetics still include note of caution:

> Epigenetic regulation—biological mechanisms that influence the expression of genes and which may be influenced by the cellular environment, over different time scales from seconds to minutes to hours, days, and years and perhaps (more controversially) across generations, and with different degrees of reversibility. Biological mechanisms (e.g., DNA methylation) that affect gene expression without changing DNA sequence. These processes may be involved in long-term developmental changes in gene expression (Thomas et al., 2015, p. 17).

It is important for those of us who are not experts in the field of biology or genetics (which the author is certainly not) to bear in mind the perspective of those who approach the topic with more caution. In addition to the cautious tone taken by Thomas above Cecil, Smith, Walton, Mill, McCroy, and Viding (2016) reviewed the evidence for epigenetic signatures for abuse and neglect, arguing that further replication will be needed before firm conclusions could be drawn. More recently still Scorza et al. (2018) have noted that despite a good level of evidence for intergenerational transmission in animal research there is still a need for more research in humans before we can say for certain that there is good evidence for these pathways impacting on areas such as disadvantage in human populations.

Nevertheless, other writers indicate a growing consensus in the research community that the changes that are made to methyl groups and histones can be passed from one generation to another and that they can contribute to developmental processes (Rutter & Pickles, 2016).

In 'Darwinian' evolution the organism gains an advantage because there is a random mutation in the genetic code itself. Subsequent generations will also inherit that altered gene. If the mutation is adaptive and gives them an advantage they will prosper. In the new paradigm life experiences programme how the genetic information is used by the body and are set as changes to DNAm. Through this mechanism the effects of these life experiences can also be inherited along with the DNA itself. "You can inherit something beyond the DNA sequence. That's where the real excitement of genetics is now" (Goldberg, Allis, & Bernstein, 2007, quoting Watson, 2003).

One of the most convincing examples of intergenerational effects is that of the Dutch Winter Hunger, or Dutch Famine.

The Dutch Famine

Many authors writing about the development of epigenetics highlight the significance of research into the Dutch famine or Dutch Winter Hunger (Carey, 2011; Spector, 2012; Rutter & Pickles, 2016; Scorza et. al., 2018). The situation in Holland in the winter of 1944-45 provided some unique conditions that have allowed researchers to investigate epigenetic processes (Heijmans et al., 2008). German restrictions on food for the Dutch population created a famine, while at the same time normal health records, food rations and health care were all maintained. Researchers were subsequently able to pinpoint individuals 6 or 7 decades later who had been conceived during this time, with detailed knowledge of the mother's health, diet and birth details. The research identified lower methylation in specific parts of the DNA sequence in these individuals. This was some of the first clear evidence that the impact of the harsh environmental conditions could be seen in the microgenetic makeup of an individual and is highlighted as some of the most significant research in the area (Rutter & Pickles, 2016). The impact adversity has on development has been explored in relation to other major events in history as well. The intergenerational impact of the Holocaust being a key example.

The Holocaust and intergenerational trauma

A large body of work has developed following extensive investigations into the experiences of Holocaust survivors and their offspring (Shmotkin et al., 2011; Kellermann, 2013; Yehuda et al., 2008; Yehuda et al., 2014; Yehuda et al., 2016).

A study of 211 adult offspring of Holocaust survivors (Yehuda et al., 2008) identified that a higher prevalence of PTSD, mood and anxiety disorders and substance abuse disorders were found in the offspring of survivors than demographically comparable Jewish controls. The investigation also identified that maternal PTSD made a greater contribution to transgenerational transmission than paternal PTSD. By 2014 Yehuda and colleagues (2014) had identified the first evidence of alteration to specific genes in the form of methylation associated with this inheritance. Later research (Yahuda et al., 2016) has shown that the epigenetic effects are present in offspring who were conceived after the trauma took place.

Shmotkin et al. (2011) emphasize that it is that is not the trauma event that is transmitted, but the impact the event has on the person experiencing it. A wide range of factors moderate the degree to which the effects of a trauma are transmitted and can protect individuals. These include the quality of the marital relationships, the existence of wider support systems for the individual after the trauma and the use of defense mechanisms to 'isolate the effects of the Holocaust from crucial aspects of their functioning" (Shmotkin, 2011, p. 10). In this way it is argued that resiliency could be transmitted through the same epigenetic processes. And although trauma-based transmission can take place from children of Holocaust survivors to the grandchildren of survivors so can resilience to trauma.

Ammaniti and Gallese (2014) describe the epigenetic transmission of resilience and intergenerational responses to stress. Drawing on a wide range of literature including animal studies they note that research is concluding that "there is no significant main effect of genes, a marginally significant effect of environment, but a relevant significant effect of the G x E interaction" (p. 170).

Addiction

Often seen as an attempt to cope with adversity substance abuse and addiction is one of the areas that has drawn some interest. Cecil, Walton and Viding (2016) review what is known about epigenetic mechanisms in relation to addiction. They acknowledge that within the current literature on addiction most of the research has been based on animal studies rather than human participants. Nevertheless, they are able to conclude that there is 'tentative evidence for intergenerational transmission of DNAm patterns implicated in addiction'. They note a number of cautions however, including that it is difficult to conclude with certainty yet that DNAm is causally linked to addiction without longitudinal studies. These have not been set up in this particular area as yet.

So, there is now evidence from a range of different sources that epigenetic processes seem to play a significant role in responses to adversity. Key psychological responses such as resilience and addiction are being linked to DNAm processes and there is growing evidence that this is inherited from one generation to the next.

The educational context: child development, learning and emotional wellbeing

As knowledge of epigenetics grows what other areas are being explored that are relevant to children's development, learning and emotional wellbeing?

Authors in pediatric journals have started to highlight the importance of epigenetic processes in young children's development. Williams and Drake (2015) write that

> There has been much interest in recent years in the role of epigenetic modifications in early life programming. Epigenetic modifications lead to changes in gene expression that are not explained by changes in DNA sequence, and during normal development, key developmental stages are characterized by epigenetic modifications that have the potential to be altered/disrupted by environmental cues (p. 1060).

The perspective taken is that those concerned with child development should know and understand epigenetic processes, because the science is robust enough for professionals to have confidence that developmental pathways are influenced significantly by epigenetic processes.

In terms of learning and learning difficulties Smith (2011, p. 356) reviewed the research looking at language and learning disorders and noted that there are perhaps a surprisingly a small number of genes that seem to be involved in early development of these skills, particularly the process of how growing neurons migrate within the cortex to their specialist areas. Smith notes that rather than specific genes for specific learning difficulties

> most of these candidate genes have been associated with several learning and language phenotypes, suggesting that they facilitate learning processes which are basic to learning reading and language……effects are seen for several genes that primarily affect autism or language but have also shown effects on reading.

So, genes don't affect single areas of learning difficulty in a specific way. The effects of gene expression happen across broad areas of learning and development.

As well as this non-specific gene effect Smith goes on to comment that although a few genes have been found that might play a role in overlapping areas of developmental difficulty

> ..very few coding mutations have been reported to account for their influence on these disorders. This has led to the hypothesis that mutations affecting reading and related disorders are likely to be in regulatory regions….[and that]… epigenetic controls of gene expression have been found that affect developmental learning disorders (2011, p. 356).

In short there are very few areas of the gene sequence itself that have been associated with learning or developmental difficulties, and where they have been found they are non-specific in their action. What is proving to be much more likely is that the epigenetic processes that regulate how groups of genes are used in growth and development are what make the differences between human beings. Language and learning difficulties are unlikely to have a specific genetic cause, but much more likely to result from epigenetic processes.

Finally, an example related to emotional wellbeing. Goodyer (2015) identifies that twin studies would typically estimate that overall genetic heritability of 'depressive symptoms' would be about 35%. This suggests that children's wellbeing is to a large degree determined by their genes. However, as with many other areas, molecular genetics (GWAS studies) has not replicated this figure. Of the GWAS studies in this area Goodyer writes that "Overall the findings do not support a strong role for genetic factors…implicating the importance of parent-child relationships" (p. 1065). Again, the problem of missing heritability indicates that if specific genes cannot be found through GWAS studies for 'depressive symptoms' then twin studies have effectively overestimated the variation that can be attributed to genes. Epigenetic processes provide the solution to this gap.

Across a spectrum of child development, learning and emotional wellbeing epigenetics is being highlighted as a crucial process to consider. What are the wider implications for the paradigm shift? How could it affect our understanding of how we effect change in our lives?

The inheritance of acquired characteristics: A new model of biological inheritance?

Through epigenetics a new model is emerging, and it offers a radically different understanding to our existing notions of Darwinain or neo-Darwinian evolution. The new conceptualization is that acquired characteristics can be inherited. This is not a completely new perspective. Scorza et al. (2018) note that the work in this area is akin to 'Reviving Lamarck'. Lamarck, a contemporary of Darwin, developed a theory of biological inheritance that was very different to Darwin's, whereby efforts of a creature to change its life could be inherited. According to Spector (2012) the most quoted example of Lamarckian evolution is that given by Lamarck about giraffes. The giraffe strives to reach food on higher and higher branches. The giraffe could, through effort, elongate its neck and then would pass on a longer neck to subsequent generations. Jablovka and Lamb (1995) discussed early epigenetic thinking in relation to Lamarckian evolution, highlighting the significance of the inheritance of acquired characteristics. Although discounted by the scientific community at the time strands of Lamarck's thinking not hold true and have the potential to radically alter the way we might think about how we effect change in our life and what the implications of this are. In a Lamarckian or neo-Lamarkian world the changes you make in your lifetime can improve your life and are passed to your children, and their children, through epigenetic processes. Could being in an environment where developing good language skills and learning to read make it easier for your children and your grandchildren to learn to read in some way even before the impact of their own childhood and education is considered? Currently an answer to this question would be little more than speculation, but findings are suggesting that such mechanisms might exist.

Implications for psychologists in education

Although more research in humans is clearly needed, (Scorza et al., 2018), the implications of the emerging understanding of epigenetic processes on learning, education and development could be profound. If the things we do in our life can affect our own epigenetic map and then that map can be transferred to our children and their children, there could be some very different ways of thinking about some aspects of educational psychology. Epigenetics could help us understand some long-standing conundrums. For example, a number of authors (Thomas et al., 2015; Brooke & Larsen, 2014, and Plonka, 2016) highlight that although twin studies seem to suggest that IQ is highly heritable the rise in IQ scores across generations (the 'Flynn effect') cannot be explained by changes in the genetic code as the pace of change is simply too fast. Epigenetic processes are being suggested as the solution to solving this problem (Greiffenstein, 2011), and although the evidence may only be emerging it has the potential to re-shape our thinking about this question fundamentally.

The approach someone takes to parenting might affect their child's resilience to stress and in turn change the epigenetic map for their generation and the subsequent generation. It is also conceivable that efforts by society, schools and teachers to increase language acquisition and literacy skills would also create an adaptive pattern whereby subsequent generations benefit from the efforts that were made at the level of biological inheritance.

Perhaps the most significant potential implication lies in the way we think about equality in society and the challenge that epigenetics brings to notions of biological determinism. To what extent are differences between us a result of things that we can change and effect and to what extent they are fixed? Epigenetics has the potential to fundamentally change how educators and society think about variation across the population. As Lewontin wrote in 1976

> The idea that inequalities are a structural element of our social organization is not a popular one and not surprisingly is regarded with hostility by the governmental, educational and information-producing agencies of our society. The alternative, which has proved more palatable and, of course, more serviceable, is that our society is pretty much as fair as any society can be and that the inequalities we observe are the irreducible differences resulting from basic biological difference between people. This is, in effect, the ideology of biological determinism (p. 6).

However, if we can change the biological inheritance we pass on by living differently, if genetics alone provides only a small explanation of the variation we see between individuals and what might explain more are the environmentally influenced mechanisms, such as DNAm, then many assumptions underpinning our collective thinking about development, education and psychology will need to change.

References

Ammaniti, M., & Gallese, V. (2014). *The birth of intersubjectivity. Psychodynamics, Neurobiology and the self*. New York: Norton.

Blum, K., Chen, A. L., Chen, T. J., Bowirrat, A., Downs, B. W., Waite, R. L., ... & Rhoades, P. (2009). Genes and Happiness. *Gene Therapy and Molecular Biology*, *13*, 91-129.

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical Twin Studies and Beyond. *Nature Reviews: Genetics*, *3*, 872-882. https://doi.org/10.1038/nrg932

Brooke, J., & Larsen, C. (2014). The Nurture of Nature: Genetics, Epigenetics, and Environment in Human Biohistory. *American Historical Review*, *119*(5), 1500-1513. https://doi.org/10.1093/ahr/119.5.1500

Carey, N. (2011). *The Epigenetics Revolution*. London: Icon Books.

Cecil, C., Smith, R., Walton, E., Mill, E., McCroy, E., & Viding, E. (2016). Epigenetic signatures of child abuse and neglect: Implications for psychiatric vulnerability. *Journal of Psychiatric Research*, *83*, 184-194. https://doi.org/10.1016/j.jpsychires.2016.09.010

Cecil, C., Walton, E., & Viding, E. (2016). Epigenetics of Addiction: Current Knowledge, Challenges and Future Directions. *Journal of studies on alcohol and drugs*, *77*(5), 688-691. https://doi.org/10.15288/jsad.2016.77.688

Chang, C., Zhang, K., Veluchamy, A., Hébert, H. L., Looker, H. C., Colhoun, H. M., ... & Meng, W. (2016). A Genome-Wide Association Study Provides New Evidence That CACNA1C Gene is Associated With Diabetic Cataract. *Investigative Ophthalmology and Visual Science*, *57*(4), 2246-2250. https://doi.org/10.1167/iovs.16-19332

Craddock, N. (2013). Genome-wide association studies: what a psychiatrist needs to know. *Advances in Psychiatric treatment*, *19*, 82–88. https://doi.org/10.1192/apt.bp.110.007906

Goldberg, A., Allis, D., & Bernstein, E. (2007). Epigenetics: A landscape takes shape. *Cell*, *128*, 635-638. https://doi.org/10.1016/j.cell.2007.02.006

Goodyer, I. (2015). Genes, environments and depressions in young people. *Archives of Disease in Childhood*, *100*, 1064-1069. https://doi.org/10.1136/archdischild-2014-306936

Greiffenstein, M. F. (2011). Secular IQ increases by epigenesist? The hypothesis of cognitive genotype optimisation. *Psychological Reports*, *109*(2), 353-366. https://doi.org/10.2466/03.04.10.19.PR0.109.5.353-366

Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., ... & Lumey, L. H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proceedings of the National Academy of Sciences*, *105*(44), 17046-17049. https://doi.org/10.1073/pnas.0806560105

Jablovka, E., & Lamb, M. (1995). *Epigenetic Inheritance and Evolution. The Lamarckian Dimension*. Oxford: Oxford University Press.

Kellermann, N. (2013). Epigenetic Transmission of Holocaust Trauma: Can Nightmares Be Inherited? *Israeli Journal of Psychiatry Related Sciences*, *50*(1), 33-43.

Lewontin, R. (1976). The Fallacy of Biological Determinism. There's no good evidence suggesting that IQ is inherited. *The Sciences*, *16*(2), 6-10.

Lock, M., & Palsson, G. (2016). *Can science resolve the nature/nurture debate?*. Cambridge. Polity Press.

Plonka, B. (2016). Nature or Nurture – Will Epigenomics Solve the Dilema?. *Studia Humana*, 5(2), 13-36. https://doi.org/10.1515/sh-2016-0007

Rutter, M., & Pickles, A. (2016). Annual Research Review: Threats to the validity of child psychiatry and psychology. *Journal of Child Psychology and Psychiatry*, *57*(3), 398-416. https://doi.org/10.1111/jcpp.12461

Scorza, P., Duarte, C. S., Hipwell, A. E., Posner, J., Ortin, A., Canino, G., ... & Program Collaborators for Environmental influences on Child Health Outcomes. (2018). Research Review: Intergenerational transmission of disadvantage: epigenetics and parents' childhoods as the first exposure. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.12877

Shmotkin, D., Shrira, A., Goldberg, M., & Yuval, P. (2011). Resilience and Vulnerability Among Aging Holocaust Survivors and Their Families: An Intergenerational Overview. *Journal of Intergenerational Relationships*, *9*, 7-21. https://doi.org/10.1080/15350770.2011.544202

Slatkin, M. (2009). Epigenetic Inheritance and the Missing Heritability Problem. *Genetics*, *182*(3), 845-850. https://doi.org/10.1534/genetics.109.102798

Smith, S. (2011). Approach to epigenetic analysis in language disorders. *Journal of Neurodevelopmental Disorders*, *3*, 356-364. https://doi.org/10.1007/s11689-011-9099-y

Spector, T. (2012). *Identically different. Why you can change your genes*. London: Phoenix.

Thomas, M., Kovas, Y., Meaburn, E., & Tolmie. A. (2015). What can the study of genetics offer to educators?. *Mind, Brain and Education*, *9*(2), 72-80. https://doi.org/10.1111/mbe.12077

Watson, J. (2003). Celebrating the Genetic Jubilee: A Conversation with James D. Watson. *Scientific American*, *288*, pp 66-69.

Williams, T., & Drake, A. (2015). What a general paediatrician needs to know about early life programming. *Archives of Disease in Childhood*, *100*, 1058-1063. https://doi.org/10.1136/archdischild-2014-307958

Yehuda, R., Bell, A., Bierer, L., & Schmeidler, J. (2008). Maternal, not paternal, PTSD is related to increased risk for PTSD in offspring of Holocaust survivors. *Journal of Psychiatric Research*, *42*, 1104-1111. https://doi.org/10.1016/j.jpsychires.2008.01.002

Yehuda, R., Daskalakis, N. P., Bierer, L. M., Bader, H. N., Klengel, T., Holsboer, F., … & Binder, E. B. (2016). Holocaust exposure induced intergenerational effects on FKBP5 methylation. *Biological psychiatry*, *80*(5), 372-380. https://doi.org/10.1016/j.biopsych.2015.08.005

Yehuda, R., Daskalakis, N. P., Lehrner, A., Desarnaud, F., Bader, H. N., Makotkine, I., ... & Meaney, M. J. (2014). Influences of maternal and paternal PTSD on epigenetic regulation of the glucocorticoid receptor gene in Holocaust survivor offspring. *American Journal of Psychiatry*, *171*(8), 872-880. https://doi.org/10.1176/appi.ajp.2014.13121571

Zuk, O., Hechtera, E., Sunyaeva, S., & Landera, E. (2012). The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences*, *109*(4), 1193–1198. doi: 10.1073/pnas.1119675109

Ben Hayes, PhD, Academic & Professional Tutor, DECPsy & Senior Educational Psychologist, University College of London**.**

Contact address: Ben Hayes, University College of London, Gower Street, WC1E 6BT London, United Kingdom. Email: b.hayes@ucl.ac.uk

Instructions for authors, subscriptions and further details:

http://ijep.hipatiapress.com

Education Science Students' Statistics Anxiety: Developing and Analyzing a Scale for Measuring their Worry, Avoidance, and Emotionality Cognitions

Günter Faber[1], Heike Drexler[1], Alexander Stappert[2] & Joana Eichhorn[1]

1) Leibniz University Hannover, Germany
2) University of Vechta, Germany

Date of publication: October 24th, 2018
Edition period: October 2018 – February 2019

To cite this article: Faber, G., Drexler, H., Stappert, A., & Eichhorn, J. (2018). Education science students' statistics anxiety: Developing and analyzing a scale for measuring their worry, avoidance, and emotionality cognitions. *International Journal of Educational Psychology*, *7*(3), 248-285. doi: 10.17583/ijep.2018.2872

To link this article: http://dx.doi.org/10.17583/ijep.2018.2872

IJEP – International Journal of Educational Psychology, Vol. 7 No. 3 October 2018 pp. 248-285

Education Science Students' Statistics Anxiety: Developing and Analyzing a Scale for Measuring Their Worry, Avoidance, and Emotionality Cognitions

Günter Faber, Heike Drexler, Joana Eichhorn
Leibniz University Hannover

Alexander Stappert
University of Vechta

Abstract

Current instruments for assessing university students' statistics anxiety prevailingly emphasize the affective construct component. In order to unfold the construct in a more exhaustive and differentiated manner, a scale for measuring university students' worry, avoidance, and emotionality cognitions was developed. In two samples of education science majors the present pilot study aimed at analyzing the scale's psychometric properties and at gaining preliminary validation results. Principal component analyses led to the formation of a unidimensional scale which appeared to be sufficiently reliable. Its relations to domain-specific self-belief and background variables turned out as theoretically expected – thus, for the time being the scale should claim criterion validity.

Keywords: statistics anxiety; worry; avoidance, and emotionality cognitions; scale development; psychometric analysis; preliminary validation results.

2018 Hipatia Press
ISSN: 2014-3591
DOI: 10.17583/ijep.2018.2872

Hipatia Press
www.hipatiapress.com

IJEP – International Journal of Educational Psychology, Vol. 7 No. 3 October 2018 pp. 248-285

La Ansiedad con la Estadística en Estudiantes de Ciencias de la Educación: Desarrollo y Análisis de una Escala para Medir su Preocupación, Evasión y Cogniciones Emocionales

Günter Faber, Heike Drexler, Joana Eichhorn
Leibniz University Hannover

Alexander Stappert
University of Vechta

Resumen

Los instrumentos corrientes para evaluar las estadísticas sobre la ansiedad de estudiantes en la universidad enfatizan predominantemente el componente del constructo afectivo. Para revelar el constructo de forma más exhaustiva y diferenciada, se desarrolló una escala para medir la preocupación, evasión y emocionalidad cognitiva de los estudiantes. En dos muestras de carreras de ciencias de la educación, el estudio piloto presente ha tenido por objetivo analizar las propiedades psicométricas de la escala y la obtención de resultados preliminares de validación. Los análisis del componente principal condujeron a la formación de una escala unidimensional que apareció ser suficientemente fiable. Su relación con la seguridad en uno mismo en ámbitos específicos y las variables de fondo resultado siendo la teóricamente esperada - así, por el momento la escala debería reclamar la validez de criterio.

Palabras clave: estadísticas de ansiedad; preocupación; evasión, y emocionalidad cognitiva; desarrollo de una escala; análisis psicométrico; resultados preliminares de validación.

2018 Hipatia Press
ISSN: 2014-3591
DOI: : 10.17583/ijep.2018.2872

Hipatia Press
www.hipatiapress.com

Quite a number of undergraduate and graduate students of the social sciences, education, psychology and business appear to struggle with statistics (Onwuegbuzie & Wilson, 2003; Zeidner, 1991). When dealing with the requirements of quantitative method courses which commonly are compulsory for earning their degree, these students mostly suffer from strong failure expectations and frequently experience feelings of apprehension and personal threat. As a result, they are at risk to develop and maintain a heightened level of anxiety in the face of statistical analyses – in particular, when being confronted with statistical tasks of data gathering, processing, and interpreting (Cruise, Cash, & Bolton, 1985).

Structurally, students' emerging statistics anxiety has to be considered a multidimensional construct reflecting the complex interplay of several cognitive, motivational, and physiological components (Rost & Schermer, 1989). Based on empirical findings of test anxiety research it can be defined as a domain-specific form of performance or evaluation anxiety which manifests as repeatedly occurring worry cognitions, task-irrelevant and interfering thoughts, marked states of emotional tension and physiological arousal (Zeidner, 1991). The worry component of test anxiety refers to the students' mental anticipation of failure and its negative consequences, whereas the emotionality component refers to their feelings of tenseness, nervousness or distress, and the physiological component refers to their perceptions of bodily symptoms. Over the past decades, there has been ample evidence for these components being distinguishable but mutually reinforcing (Deffenbacher, 1980; Hodapp & Benson, 1997; Kieffer & Reese, 2009; Sarason, 1984). In most cases they could be demonstrated to negatively affect the students' learning process and achievement outcomes. However, the worry component generally turned out to most strongly predict academic performance or test results (Cassady & Johnson, 2002; von der Embse, Jester, Roy, & Post, 2018). This debilitating effect of worry cognitions appeared to be mainly caused by their strongly biased and task-irrelevant mode of information processing (Schwarzer, 1996; Zeidner, 1998).

Based on this conclusive body of evidence a theoretically and methodologically sound framework for the statistics anxiety construct should implicitly consider its cognitive, emotional, and physiological components. In particular, it is essential that it addresses the issue of relevant worry cognitions as they are closely linked to the debilitating effects of statistics anxiety on statistical learning and performance. To date, however, most empirical analyses in the field have emphasized the emotional and physiological components of statistics anxiety (Cruise et al., 1985). They define the construct as feelings of anxiety or as habitual anxiety in the face of statistically loaded situations (Macher, Paechter, Papousek, & Ruggeri, 2012; Onwuegbuzie & Wilson, 2003). Admittedly, the measurement items used in these empirical analyses cover a wide range of statistical tasks that students typically encounter in everyday study or the context of their course. These items can be empirically assigned to various situation- or task-specific dimensions. Thus, for instance, factor analyses of the task- and course-related items of Zeidner's (1991) Statistics Anxiety Inventory led to the development of a content- and a test-specific subscale. Likewise, factor analyses of the widely used Statistics Anxiety Rating Scale (Cruise et al., 1985) and the conceptually related Statistical Anxiety Scale (Vigil-Colet, Lorenzo-Seva, & Condon, 2008) provided separate subcomponents concerning the students' interpretation and test anxiety, their fear of asking for help and fear of statistics teachers. Similarly, the Statistics Anxiety Measure (Earp, 2007) and the Statistics Comprehensive Anxiety Response Evaluation (Griffith et al., 2014) revealed some distinct task- or situation-specific subcomponents referring to statistically relevant course requirements and situations. In summary, these approaches to measuring the construct of statistics anxiety undoubtedly represent typically anxiety-evoking task features and test situations in a most elaborate way. However, with the exception of the Statistics Anxiety Measure (Earp, 2007) and the short research scale Hong and Karstensson (2002) used, both of which include single worry items, it is notable that all other instruments fail to integrate students' cognitive, emotional, and physiological anxiety reactions, in particular with regard to the critical worry component.

In contrast, a concurrently operating research line in the test anxiety field had already decomposed the statistics anxiety construct and assessed the students' worry and emotionality responses separately. However, in most cases a composite score including both components was used because of both components' high interrelation (Benson, Bandalos, & Hutchinson, 1994; Finney & Schraw, 2003; González, Rodrígez, Faílde, & Carrera, 2016;

Hong & Carstensson, 2002). That way, albeit merely having total anxiety scores available, the interpretation of students' responses explicitly allowed for a traceable cognitive perspective. As this research line essentially contributes to refine the statistics anxiety construct with respect to its motivationally operating components, its task- or situation-specific references appear less differentiated. That is, in all studies the items for assessing statistics anxiety referred exclusively to the taking of a statistical test or exam. This contextual limitation, hence, should challenge the representativity or content validity of the worry and emotionality measures, because their scores account only for a particular part of the relevant learning setting (Haynes, Richard, & Kubany, 1995).

Moreover, another issue crucial to the conceptualization of the statistics anxiety construct refers to the role of the students' avoidance tendencies. Already in the very first beginning of empirical test anxiety research, Mandler and Sarason (1952) posited anxiety responses to manifest as "implicit attempts at leaving the test situation" (p. 166). Subsequently, empirical findings lent support for this assumption and yielded sound evidence for students' avoidance cognitions being substantially related to their anxiety responses (Blankenstein, Flett, & Watson, 1992; Galassi, Frierson, & Sharer, 1981; Hagtvet & Benson, 1997; Skaalvik, 1997). In particular, the analyses of Elliot and McGregor (1999), Pekrun, Elliot and Maier (2009), and Putwain and Symes (2012) demonstrated clearly that students with heightened avoidance orientations reported a higher extent of worry cognitions and lower scores on subsequent exam performance. Heretofore, only few conceptualizations of the test anxiety construct had addressed this issue and claimed the students' escape or avoidance cognitions to constitute an essential part of their worries and to represent an important factor to elicit interfering, task-irrelevant thoughts (Pekrun, Goetz, Perry, Kramer, Hochstadt, & Molfenter, 2004; Schwarzer & Quast, 1985). Accordingly, further research should develop appropriate measurements being designed not only to assess the students' worries about threatening failure outcomes, but also to inquire their thoughts to preferably avoid getting involved with threatening tasks or situations. Currently available questionnaires for measuring students' statistics anxiety either do not consider their avoidance cognitions at all (Griffith et al., 2014; Onwuegbuzie

& Wilson, 2003; Vigil-Colet et al., 2008) or include just a single item assessing avoidance behavior (Earp, 2007). Research in the field should refine its conceptualization of students' worry responses and develop instruments that explicitly capture avoidance cognitions with respect to statistically loaded tasks and situations. That way, an important step to refine the substantive and structural stage of construct validation would be done (Benson, 1998).

Approaching refined measurement

To overcome the conceptual limitations of current instruments as a very first attempt a new scale to assess university students' statistics anxiety was developed. Nevertheless, it should adopt the particular strengths of existing instruments. Thus, it was assigned to approach a refined measure of the construct by meeting the following criteria: Its items should (1) specifically take into account the students' anxiety responses in a most differentiable way and, hence, consider their worry and avoidance cognitions as well as their emotional reactions. Thereby, its items should embed the various anxiety reactions (2) into a representative range of statistically loaded task features and course situations the students would typically encounter.

The construction of this scale for measuring university students' "Worry, Avoidance, and Emotionality Cognitions Encountering Statistical Demands" (WAESTA) largely followed the procedure of facet theory using a mapping sentence (Guttman & Greenbaum, 1998). This mapping sentence served as a heuristic device to cover all major facets of the construct and, thus, to achieve sufficient content validity (Edmundson, Koch, & Silverman, 1993). A pool of eligible items was drawn up using this constitutive mapping sentence which included conceptually relevant anxiety components, situational references, and intended response categories (Zeidner, 1998). In particular, each item to represent the statistics anxiety domain was specified with respect to four key facets: a relevant reaction facet referring to the worry, avoidance, and emotionality component, and three contextual facets referring to the (1) outcome in a statistics exam, the (2) individual learning of statistical procedures and handling of statistical demands, and the (3) public mastering of statistical content. Furthermore, an additional range facet defined the response categories to assess the students' perceived magnitude of individual anxiety reactions. As seemingly appropriate response range a four-point format was decided – in order to avoid artificial complexities in the respondents' decision making but instead to ensure a cognitive-motivationally realistic as well as just manageable number of rating references. This four-facet mapping sentence was used as a cross-classification template to systematically operationalize the statistics anxiety construct as it allowed to operationalize the various elements of facets in a most differentiated manner (Hox, 1997). That way, a final scale version with 17 four-point Likert-type rating items was built (Table 1).

Table 1
WAESTA items assessing relevant anxiety reactions to statistically loaded task features and course situations

Test Anxiety Component	Course Exam Outcome	Understanding Explanation Application	Oral Task Presentation Explanation
Worry	01, 14, 16	05, 08	10, 11, 12
Avoidance	03	06, 17	04
Emotionality		02, 07, 09, 13	15

Sample item and response range:

I would hardly be able to present a report on statistical research findings adequately.

Does not apply	1	2	3	4	Applies in full

All items concerned a mentally imaginable situation the students should easily manage to anticipate. Eight items referred to the students' worries about their potentially expected failure to master the course exam and to cope with several statistical requirements. Four items concerned their cognitions to preferably avoid the statistics course and particular statistical demands. Five items are related to their emotional tension when being confronted with a certain statistical task (Appendix).

As statistically indicated task requirements for the understanding of course contents, the interpretation of quantitative research results as well as the application of statistical formulas and procedures were considered. Likewise, the oral presentation and explanation of statistical content in the public course situation was included. To warrant conceptual clarity, the avoidance items should neither tap the students' avoidance reactions by suppressing or substituting individually occurring threat cognitions (Williams, 2015), nor should they refer to the students' actual avoiding strategies to cope with disliked or threatening academic events (Onwuegbuzie, 2004). Rather they should operationalize the students' mentally processed avoidance thoughts or even escape illusions before or during statistical task completion – and, thus, might indicate a specific subcomponent of worry cognitions. Moreover, the component of physiological symptoms or bodily tensions was not explicitly addressed. Instead it was thought to be indirectly inferred from the emotionality items. Conceptually, this restriction seemed to be justifiable, since the students' perceived affective state should always reflect their actual physiological arousal (Zeidner, 1998).

Validation framework and objectives

As test anxiety is assumed to be a multifaceted construct (Zeidner, 1998), first of all, the factor structure of the WAESTA scale should be analyzed. As the WAESTA items were theoretically designated to represent each the worry, avoidance, and emotionality component of the statistics anxiety construct, a clear three factor solution appeared to be expectable, at best. However, relevant research findings in the test anxiety field had demonstrated these components being substantially correlated (Deffenbacher, 1980; Cassady & Johnson, 2002; Hodapp & Benson, 1997; Hong & Karstensson, 2002; Sarason, 1984). Furthermore, in certain research contexts dealing with school students' domain- or subject-specific test anxieties, all worry, avoidance, and emotionally items repeatedly loaded on one common anxiety factor (Faber, 1995, 2012b). Therefore, an accurate prediction of the scale's ultimate factor structure seemed difficult. Rather the present study should explore the scale's underlying structure in a most tentative way – and should, thus, take into account three alternatives: a three factor solution separating the worry, avoidance, and emotionality components, a two factor solution with the worry and avoidance items loading on a first factor and the emotionality items loading on a second factor, and a one factor solution subsuming all items. Presuming the avoidance component to represent a specific worry element, the two factor solution could definitely reveal a reasonable perspective, in particular. With the reservation of this initially performed analysis, the final version of the WAESTA scale should be determined and its psychometric properties examined.

As relevant cognitive-motivational constructs academic competence and control beliefs were assessed (Schunk & Zimmerman, 2006) which are well proven to regulate students' engagement and learning approach in the long term. In particular, they essentially affect the students' anxiety experience. As unfavorable competence beliefs usually come along with increased expectancies of failure, they will provoke a strong sense of personal threat and, thus, lead to an individually heightened level of anxiety. As the students, likewise, are not (or not anymore) able to realize individually feasible perspectives to prevent a certain failure outcome, they will develop reduced control beliefs which all the more strengthen their feelings of threat and anxiety.

There is sound evidence that domain-specific academic competence beliefs or self-concepts substantially predict the individually existing magnitude of test anxiety (Ahmed, Minnaert, Kuyper, & van der Werf, 2012; Goetz, Pekrun, Hall, & Haag, 2006). Correspondingly, in the statistics domain the crucial role of self-concepts had been well established. In most cases, high-anxious students reported a lowered self-concept of own mathematics or statistical competencies (Bandalos et al., 1995; Benson, 1989; González et al., 2016; Macher et al., 2012; Williams, 2014; Zeidner, 1991). Therefore, it should be assumed the WAESTA scale scores to correlate negatively and substantially with the students' mathematics self-concept. As well, to sufficiently clarify the domain-specificity of the WAESTA scale, its relation to the students' verbal self-concept should be concurrently analyzed. According to the multidimensional feature of academic self-beliefs (Green, Martin, & Marsh, 2007), research in the field could consistently demonstrate the students' mathematics anxiety being substantially related to their performance and motivation in the mathematics but not in the verbal domain (Goetz, Frenzel, Pekrun, Hall, & Lüdtke, 2007; Gogol, Brunner, Preckel, Goetz, & Martin, 2016). From this validation perspective, the WAESTA scale should claim preliminary subject-specificity if its correlation with the verbal self-concept variable would turn out to be distinctly weaker than with the mathematics self-concept variable.

Besides, the students' control beliefs largely manifest as implicit theories or mindsets which may stress an entity view of more or less fixed and unchangeable abilities – or an incremental view of more or less modifiable and changeable abilities (Dweck & Leggett, 1988). They could be demonstrated to significantly affect the students' motivational orientations, learning strategies, and, eventually, their task performance (Blackwell, Trzesniewski, & Dweck, 2007; Burnette, O'Boyle, VanEpps, Pollack, & Finkel, 2013; Cury, DaFonseca, Zahn, & Elliot, 2008). These implicit theories principally might not only concern an individual's cognitive ability but also might emerge in a domain-specific manner and refer to the perceived malleability of certain skills or competencies (Dweck & Molden, 2005). Consequently, they should also play a motivationally crucial role in the students' learning of statistics. With respect to the statistics domain, an entity view of own competencies would diminish or even suspend any control perspective. Unfortunately, previous studies in the field had seldom analyzed the role of implicit theories. If at all, they had referred to the students' general ability beliefs, but not to specific beliefs about statistical competencies (Zonnefeld, 2015) or had only considered the students' beliefs to master statistical demands through strategy use and effortful behavior (Schutz, Drogosz, White, & Distefano, 1998) – thus, reflecting an incremental view of learning approach. However, these learning control beliefs could be demonstrated to significantly predict course grades. Accordingly, against the background of research findings the WAESTA scale scores should be reasonably assumed to correlate positively and substantially with the students' entity view of less or not malleable statistical competence.

Furthermore, as another motivational criterion variable, the students' task values were considered. Conceptually, from an expectancy-value perspective on achievement motivation task values concern the students' perceived importance or adequacy of a certain activity to fulfill their personal needs and to attain their personal goals (Eccles & Wigfield, 2002). In particular, these task values had been demonstrated to regulate the students' motivational orientations, learning strategies, and academic choices (Wigfield, Hoa, & Klauda, 2009). Task values evidently emerge in a task- or at least domain-specific manner (Gaspard, Häfner, Parrisius, Trautwein, & Nagengast, 2017; Selkirk, Bouchey, & Eccles, 2011). Thus, they should characteristically affect the students' learning and performance in the statistics domain as well. Previous research in the field had primarily analyzed the students' perceived utility or worth of statistical knowledge and competencies as an attitudinal construct (Nolan, Beran, & Hecker, 2012) focusing on the usefulness of statistics in personal and professional contexts (Cruise et al., 1985; Dauphinee, Schau, & Stevens, 1997). The students' ratings of the worth of statistics appeared to positively correlate with their statistical achievement as well as with their learning strategies to a slight extent only (Emmioğlu & Capa-Aydin, 2012). In comparison, the relations of utility perceptions with domain-specific measures of academic self-beliefs were stronger. Students with low competence beliefs valued statistics as less important (Baloğlu, 2002; Chiesi & Primi, 2009; Dauphinee et al., 1997; Vanhoof, Kuppens, Sotos, Verschaffel, & Onghena, 2011). Similarly, students' statistics anxiety was also moderately correlated with their utility ratings – indicating those students suffering from a heightened level of statistics anxiety tendentially perceived statistics to a lesser extent as useful (Baloğlu, 2002; Chew & Dillon, 2014; Nasser, 2004; Papanastasiou & Zembylas, 2008; Papousek et al., 2012). Accordingly, the WAESTA scale scores should be assumed to correlate inversely and substantially with the students' perceived value of statistical competence.

As a relevant background variable to explain the students' statistical self-beliefs and competencies their prior mathematical learning, in particular their latest school grade had been well proven. From the perspective of self-concept development (Marsh & O'Mara, 2008), previous failure experience in mathematics will evidently lead to form low competence beliefs in the statistics domain and, eventually, contribute to strengthening the emergence of domain-specific anxiety responses. In various studies students with poor school grades in mathematics reported a heightened level of statistics anxiety (Beurze, Donders, Zielhuis, de Vegt, & Verbeek, 2013; Birenbaum & Eylath, 1994; Chiesi & Primi, 2010; Lalonde & Gardner, 1993).

Accordingly, the WAESTA scale scores should be assumed to correlate negatively, but low in magnitude with the students' mathematics grade they had last earned at school.

Method

Participants and procedure

In the present study the data of both a construction sample and a validation sample were analyzed. The construction sample consisted of 113 graduate students (n = 94 females, n = 19 males) from a German university Master's course in educational sciences (n = 80) and special education (n = 33). They all were enrolled in a compulsory course on empirical research methods. Therefore, the participation rate was sufficiently high at 82 per cent. Seventy-four of the students had already acquired elementary statistical knowledge during their first degree, whereas 39 were required to attend a course in basic descriptive and inferential statistics. Both the subgroup with and without statistical knowledge did not significantly differ with respect to gender (chi-square test, $p > .05$) and age (Mann-Whitney U-test, $p > .05$). Also, there was no significant difference of gender ratio within each subgroup (binomial test, $p > .05$).

The validation sample was thought to scrutinize the findings from the construction sample one year later. It consisted of 87 graduate students from the same Master's courses: educational sciences (n = 59) and special education (n = 28). The sample was predominantly female (n = 74). As with the construction sample all the students were enrolled on a compulsory course on empirical research methods. The participation rate was rather high at 89 per cent. Fifty of the students had acquired statistical knowledge during their first degree whereas 37 had to attend an introductory statistics course. Once again there were no subgroup differences in gender, gender ratio, or age (Stappert, 2017).

In both samples all relevant data concerning the self-belief and background variables under consideration were gathered on the course's first term. For that purpose, a questionnaire including all items to measure the students' self-concept, statistics anxiety, implicit theories, task values, and relevant background information was administered. To prevent a priming effect of the self-concept items, they were presented at the end of the questionnaire.

Both samples had missing data (5.7% and 7.3%). As they did not produce any systematic pattern in the construction (MCAR test $p = .182$) and in the validation sample (MCAR test $p = .178$), they were treated as "missing completely at random" (Little, 1988). The missing values were estimated by means of the two-step iterative expectation-maximization algorithm (Graham, 2012).

Measures

Students' academic self-concepts in mathematics and language (German) were assessed using nine six-point rating items for each subject. These items referred to the students' most recent learning experiences at school and addressed their competence beliefs with regard to meeting subject-specific demands. The wording of the items was strictly parallel. In the majority, the items originate from well proven instruments (Faber, 2012a; Möller, Streblow, Pohlmann, & Köller, 2006; Rost, Sparfeldt, & Schilling, 2007). For the purpose of this study they were adapted and phrased retrospectively. Sample item: "I tried hard in mathematics/German, but I did not perform very well." Principal component analysis (with varimax rotation) revealed a two-factor solution allowing for a clear distinction between the subject-specific self-concept facets. Hence, it was possible to build two scales for measuring the subject-specific academic self-concepts. Their reliability was most appropriate for both the mathematics and the language self-concept scale (Table 2). High scale scores indicated the students' competence beliefs being positive. According to the multifaceted feature of the construct, the self-concept variables appeared to lowly correlate in the construction sample ($r = .14$, $p > .05$) and in the validation sample ($r = -.05$, $p > .05$).

To assess students' implicit theory of statistical competencies, a short scale with five four-point rating items was administered in the construction sample. As current instruments in the field only allowed for measuring the students' implicit intelligence theory (İlhan & Cetin, 2013; Kooken, Welsh, McCoach, Johnston-Wilder, & Lee, 2016), a new scale was created. Following the recommendations of Hong, Chiu, Dweck, Lin and Wan

(1999), all items tapped an entity view of personal statistical competence. Unfortunately, due to their insufficient item-test correlation ($r_{it} < .22$) two items had been deleted. With an average item intercorrelation of Fisher's z' = .44 and an average item-test correlation of Fisher's z' = .52 the final scale's reliability appeared to be just acceptable. Sample item: "To work with statistics, you need a talent that I simply do not have." High scale scores indicated the students to perceive their statistical ability being fixed, hence as less malleable in nature. In the validation sample, a slightly revised scale with four four-point Likert items was used (Stappert, 2017). In view of sample size and item number its reliability appeared to be sufficient (Table 2).

Table 2

Descriptive statistics and reliabilities of the scales for measuring validation variables

	Items	*AM*	*SD*	z_S	z_K	α
Latest School Grade in Mathematics						
Construction Sample	1	3.06	0.91	0.76	-0.15	
Validation Sample	1	3.11	1.36	-0.50	-2.28*	
Academic Self-Concept in Mathematics						
Construction Sample	9	29.81	9.81	1.37	-2.04	.93
Validation Sample	9	31.01	11.43	-0.33	-1.84	.94
Academic Self-Concept in Language (German)						
Construction Sample	9	45.52	7.32	-3.61***	-0.01	.92
Validation Sample	9	41.45	8.53	-2.12*	-0.16	.90
Implicit Entity Theory of Statistical Competencies						
Construction Sample	3	5.43	2.02	3.59***	0.13	.70
Validation Sample	4	7.87	2.61	2.15*	-0.32	.81
Negative Instrumental Value of Statistics						
Construction Sample	5	11.06	2.89	0.41	-1.18	.72
Validation Sample	8	16.18	3.95	0.55	-0.10	.75

Table 2 (continued)

Descriptive statistics and reliabilities of the scales for measuring validation variables

Significance: $*p \leq .05$, $***p \leq .001$

AM = arithmetic mean, *SD* = standard deviation, z_S = z-standardized skewness, z_K = z-standardized kurtosis, α = internal consistency (Cronbach's coefficient alpha)

The utility value students attributed to statistical competence was measured by means of a short scale. In the case of the construction sample it consisted of five four-point rating items dealing with the perceived utility of statistics for the students' current studies and intended career. Sample item: "Statistics will not play an important role in my future professional life". With an average item intercorrelation of Fisher's z' = .40 and an average item-test correlation of Fisher's z' = .49 the scale's reliability was just acceptable (Table 2). High scale scores indicated the students to consider statistics being less important. In the validation sample, an extended version of the scale was used. It consisted of eight four-point Likert items. Its reliability was once more just acceptable (Table 2). Here again, high sum scores indicated students to perceive statistics as being less useful for their current studies and later professional development (Eichhorn, 2018).

Finally, as relevant background variable the students' most recent school grade in mathematics was inquired in both samples.

Results

Scale formation

For determining the final version of the WAESTA scale in the construction sample, first of all, descriptive item statistics were calculated. The avoidance item 04 as well as the worry item 11 showed a significant negative skew indicating most students to agree with the statements – in detail they would preferably give a presentation without any statistical content and during a presentation they would strongly hope not being asked statistical questions. Furthermore, as the analysis revealed a significant negative kurtosis score for the items 03, 06, 14, and 15, their distribution appeared to be platykurtic.

Accordingly, the students' relevant item responses denoted a heightened variance or difference among them (Table 3).

In the construction sample, the Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of sphericity demonstrated the inter-item correlations being appropriately strong (KMO = .878, BTS $p < .001$). Therefore, a principal component analysis (PCA) was conducted in order to clarify the latent scale structure. However, it could not statistically separate the three anxiety components. Neither a varimax nor an oblique rotation procedure could yield any loading pattern to separate the worry, avoidance, and emotionality items in a conceptually proper way. Rather all analyses led to a unidimensional structure (Table 3). This solution revealed sufficiently high factor loadings and could explain 43.59 per cent of extracted variance. Though for further clarification three provisional subscales representing the students' worry, emotionality, and avoidance cognitions were formed and the relations among their sum scores examined. In line with the PCA result, the subscales were strongly correlated – in particular, the worry with the emotionality scale $r = .80$ and with the avoidance scale $r = .72$, the emotionality with the avoidance scale $r = .69$ (all $p < .001$).

Consequently, all WAESTA items could be used to build the scale's final version. For its total score, neither the z-standardized scores of skewness and kurtosis nor the Shapiro Wilk W-test ($W = .988$, $df = 113$, $p = .443$) could evince any significant deviation from the normal distribution assumption. High total scores indicated the students' to report stronger worry, avoidance, and emotionality cognitions. The scale's reliability was estimated in various ways and turned out to be adequate: Its internal consistency (Cronbach's coefficient alpha) amounted to $\alpha = .92$, its split-half reliability (odd-even method using Spearman-Brown correction) to $r_{12} = .89$, and its standard error (based on coefficient alpha) was $s_e = 2.67$.

Table 3

Descriptive statistics, factor loadings and corrected item-test correlations of WAESTA items (WR = worry, AV = avoidance, EM = emotionality): Results from the construction sample

Item	AM	SD	z_S	z_K	a	r_{it}
01 WR	2.37	0.82	1.49	-0.67	.486	.436

Table 3 (continued)
Descriptive statistics, factor loadings and corrected item-test correlations of WAESTA items (WR = worry, AV = avoidance, EM = emotionality): Results from the construction sample

02 EM	2.71	0.94	-1.30	-1.69	.688	.645
03 AV	2.74	1.08	-1.33	-2.61**	.581	.515
04 AV	2.88	0.97	-2.12*	1.65	.719	.664
05 WR	2.51	0.91	-0.50	-1.68	.725	.663
06 AV	2.50	0.97	-0.21	-2.08*	.758	.697
07 EM	2.17	0.96	1.71	-1.78	.697	.620
08 WR	2.24	0.84	1.49	-0.83	.718	.670
09 EM	2.20	0.87	1.59	-1.04	.668	.583
10 WR	2.69	0.93	-1.63	-1.43	.647	.694
11 WR	2.91	0.97	-2.09*	-1.75	.740	.676
12 WR	2.26	0.80	1.52	-0.45	.734	.618
13 EM	2.96	0.93	-1.51	-0.45	.683	.493
14 WR	2.69	1.07	-0.67	-2.18*	.551	.474
15 EM	2.87	0.84	-1.30	-2.77**	.532	.637
16 WR	2.75	0.84	-1.45	-0.83	.689	.442
17 AV	2.07	0.87	1.55	-1.52	.488	.442
Total Sum Score (Items 01-17)						
WAESTA	42.66	10.20	-0.49	-1.34		

Significance: **p* ≤ .05, ***p* ≤ .01
AM = arithmetic mean, *SD* = standard deviation, z_S = z-standardized skewness, z_K = z-standardized kurtosis, *a* = factor loading, r_{it} = corrected item-test correlation

In spite of the small number of participants in the validation sample, a principal component analysis (PCA) of the WAESTA items was conducted. As their communalities ranged from h = .412 to h = .660 and the Kaiser-Meyer-Olkins measure revealed an appropriately high score (KMO = .901), this procedure appeared to be most reasonable (de Winter, Dodou, & Wieringa, 2009; MacCallum, Widaman, Zhang, & Hong, 1999). The results revealed one common factor with considerably high loadings (ranging from a = .532 to a = .804). Accordingly, the unidimensional scale feature could be fully replicated and explained 50.66 per cent of extracted variance. For its total sum score, z-standardized skewness (z_S = -0.330) and kurtosis values (z_K = -0.190) did not indicate any significant deviation from the normal distribution assumption. Here again, the scale's reliability was estimated in various ways and turned out to be adequate: its internal consistency (Cronbach's coefficient alpha) amounted to α = .94, its split-half reliability (odd-even method using Spearman-Brown correction) to r_{12} = .91, and its standard error (based on coefficient alpha) was s_e = 2.45.

Validation results

As an initial approach to analyze the external validity of the WAESTA scale, its zero-order correlations with the criterion variables under consideration were first analyzed. As the results could demonstrate (Table 4), the WAESTA scores were closely and significantly associated with the students' mathematics self-concept but not with their language self-concept. This particular finding might be considered to provide preliminary evidence that the WAESTA scale measures rather a domain-specific than a general facet of the students' test anxiety experience.

Table 4

Relations of WAESTA sum scores with background and self-belief variables (zero-order correlations): Results from the construction and the validation sample

Most Recent School Grade Mathematics	Academic Self-Concept Mathematics	Academic Self-Concept Language	Implicit Entity Theory Competence	Negative Instrumental Value
Construction Sample				
-.31***	-.43***	.09	.62***	.49***
Validation Sample				
-.22*	-.38***	-.08	.80***	.32***

Significance: *$p \leq .05$, ***$p \leq .001$

Furthermore, the scale scores were most strongly correlated with the students' entity views of own statistical competence. A heightened level of statistics anxiety came along with a deep understanding of own statistical competencies being less or even not malleable in nature. With the negative instrumental value of statistics, the WAESTA sum score correlated moderately positive. Students reporting a higher level of statistics anxiety tendentially perceived statistical competencies as less important. Finally, the relation between the WAESTA score and the most recent school grade in mathematics appeared to be positive and significant, though low in magnitude. Hence, students with a heightened level of statistics anxiety had been less successful in the mastery of mathematical demands at school.

To get more differentiated validation results, a series of regression analyses with the WAESTA scale as dependent variable were computed for both samples. As this procedure allowed for controlling the covariations among all predictor variables with respect to their empirical overlap and multicollinearity, it should help to unravel the complexity of construct relations. In particular, a sequence of regression models including an advancing number of predictor variables was consecutively tested (Table 5). In both samples the standardized residuals of WAESTA sum scores did not violate the normal distribution assumption. In each case, the Shapiro Wilk W-test could demonstrate the standardized residuals being normally distributed (construction sample: $W = 986$, $df = 113$, $p = .308$, validation sample: $W = 979$, $df = 87$, $p = .182$).

The results for regression model A clearly demonstrated the mathematics self-concept to explain the most part of anxiety variance. However, adding the entity beliefs to the regression equation in model B and C, the predictive power of the students' mathematics self-concept was reduced to a minimal and insignificant extent. Instead, the students' entity beliefs largely contributed to the WAESTA sum score. As the mathematics self-concept and the entity belief variable in both samples were substantially correlated ($r = -.51$ in the construction sample and $r = -.45$ in the validation sample), but the entity belief variable in both samples was more strongly related to the anxiety variable ($r = .63$ in the construction sample and $r = .80$ in the validation sample) – the massive decline in the self-concepts' beta weight must be seen as a result of multicollinearity. This predictive pattern occurred in both samples. Moreover, only in the construction sample the students' negative value of statistics substantially and independently explained additional variance in the WAESTA sum score. The difference between samples might be due to the fact, that the methods for assessing the value variable were not comparably formatted. Apart from this, all regression analyses demonstrated the students' statistics anxiety to be essentially and most closely predicted by their control beliefs – as reflecting their perceived malleability of individual competencies in the statistics domain.

Table 5.
Multiple regression of WAESTA sum scores on background and self-belief variables (standardized beta weights and squared multiple regression coefficients): Results from the construction and the validation sample

Model	Most Recent School Grade Mathematics	Academic Self-Concept Mathematics	Implicit Entity Theory Competence	Negative Instrumental Value	R^2
	Construction Sample				
A	.027	-.452***			.154
	Validation Sample				
A	.184	-.533***			.182
	Construction Sample				
B	-.070	-.090	.545***		.396
	Validation Sample				
B	.031	-.089	.761***		.620
	Construction Sample				
C	-.093	-.045	.452***	.286***	.465
	Validation Sample				
C	.024	-.083	.789***	-.062	.614

R^2 = adjusted multiple regression coefficient squared
Significance: *$p \leq .05$, **$p \leq .01$, ***$p \leq .001$

Finally, for further clarification of the WAESTA scores, their mean differences between both the educational science and the special education students were analyzed. As the comparison groups were small and unequal in size the nonparametric Mann-Whitney U-test for independent samples was used. In the construction sample a significantly higher level of statistics anxiety in the special education subgroup could be found ($Z = -2.314$, $p = .021$, effect size $r = 0.22$). In the validation sample, the level of statistics anxiety did not significantly differ between educational science and special education students ($Z = -0.374$, $p > .05$). However, with the small size of the validation sample in mind, this finding should be considered cautiously.

Discussion

The present study should examine the internal and external validity as well as the psychometric properties of the newly developed WAESTA scale for measuring educational science students' worry, avoidance, and emotionality cognitions in the domain of statistics learning. Conceptually, this measurement approach should integrate both the strengths of a more situation- and a more reaction-focused research line in the field. As a substantive result, this scale could be demonstrated to represent the construct in a unidimensional manner. The final scale version included all items as initially administered in both samples. Its internal consistency was most sufficient. Furthermore, its relations with self-belief and attainment variables most widely turned out as theoretically predicted. Specifically, total WAESTA score was more strongly correlated with the students' mathematics than with their language self-concept – and, thus, the scale could claim domain-specific validity. These findings correspondingly held for both the construction and the validation sample. For the time being, the WAESTA scale can be considered internally and externally valid as well as having adequate psychometric properties. Nevertheless, some results definitely deserve further attention.

In particular, the scale's underlying structure consistently appeared to be unidimensional. This finding indicates the strong empirical overlap among the worry, avoidance, and emotionality responses – and, thus, the cognitive-motivational interplay of anxiety components. Similarly, close relations had been already found elsewhere (Deffenbacher, 1980; Hodapp & Benson, 1997; Cassady & Johnson, 2002; Chin, Williams, Taylor, & Harvey, 2017;

Hong & Karstensson, 2002; Sarason, 1984), especially with respect to domain- or task-specific facets of test anxiety (Faber, 1995, 2012b). By no means, this result does challenge the need for a separate assessment of worry, avoidance, and emotionality cognitions. Rather this approach should ensure to obtain a more differentiated measuring of statistics anxiety and, thereby, should contribute to reducing the interpretation ambiguity of item responses. In that regard, it should certainly increase the scale's cognitive-motivational representativity and content validity. Likewise, the students' avoidance cognitions were found to be most closely related to their worry, but only slightly less closely related to their emotionality cognitions. Therefore, according to relevant findings (Galassi et al., 1981; Hagtvet & Benson, 1997), avoidance cognitions must be seen as an important feature within the students' anxiety experience and, thus, should have contributed to completing and refining the measuring of statistics anxiety (Putwain, 2008).

With respect to the validation results, both the correlation and regression analyses suggest, at first glance, that students' implicit entity beliefs are sufficient to explain their statistics anxiety. The entity beliefs appear to obviously play a crucial role in the prediction of statistics anxiety – as could be expected from the view of social-cognitive theories (Dweck & Leggett, 1988; Schunk & Zimmerman, 2006). However, in the students' cognitive-motivational processing, they will operate in a more complex manner. According to relevant theoretical conceptions and empirical findings (Blackwell et al., 2007; Chiesi & Primi, 2010; Emmioğlu, 2011; Onwuegbuzie, 2003; Sesé, Jiménez, Montano, & Palmer, 2015), it should be assumed that implicit beliefs actually mediate the effects of students' self-concept and their learning background on the dependent anxiety variable. The massive decline in the self-concept variable's beta weights, when adding the entity belief variable to the regression equation, apparently supports this assumption (Table 5). Indeed, within the validation framework this indirect effect cannot be adequately substantiated with correlational or regression analysis, but only with multivariate modeling method (Kline, 2011). Future research should make every effort to apply such modelling techniques in order to clarify the role of entity beliefs in the statistics domain.

Beyond the purpose of scale validation, the empirical findings concerning the students' entity beliefs might even extend the previous research in two respects: at the level of construct specificity, the measuring of entity beliefs did not refer to the perceived malleability of general cognitive abilities but enquired the perceived malleability of statistical competencies. As this entity belief variable was only very weakly correlated with language self-concept (construction sample r = .07; validation sample r = -.12) it should be considered domain-specific. Hence, for the domain of statistical learning, this particular finding appears to be in line with the recommendations of the implicit theories approach (Dweck & Molden, 2005). Accordingly, at the level of construct relations, the results allow for refining the nomological scope of the statistics anxiety framework – at least, as it refers to the type and role of self-belief variables (Bandalos et al., 1995; González et al., 2016; Onwuegbuzie & Wilson, 2003; Zeidner, 1991).

The present study undeniably suffers from some conceptual and empirical limitations. First of all, composition and size of both student samples do not allow for generalizing the empirical findings as should be required. Instead, the findings reported here might claim a sort of local validity – all the more, as their data basis referred to a certain university setting. Further analyses should necessarily remedy this problem and examine the WAESTA scale with other student samples from other educational science contexts.

Moreover, the validation framework is still lacking in several respects and should be further completed (Benson, 1998). The present study assessed students' mathematics self-concept retrospectively. As a proportion of both samples did not have any prior experience of statistics using a measure of statistical self-concept would have been misguided. However, provided that further research could include participants being most comparable in their statistical background, their self-concept in the statistics domain should be absolutely used to elaborate scale validation (González et al., 2016). Likewise, concurrent measures of the students' self-efficacy to master certain statistical tasks could help to further differentiate the scale's criterion validity (Finney & Schraw, 2003; Perepiczka, Chandler, & Becerra, 2011). Not least, an appropriate validation of the WAESTA scale will require analysis of its relations with other instruments for measuring statistics anxiety – for instance, by comparing it with the German adaptation of the STARS questionnaire (Papousek et al., 2012).

Another considerable lack of the present study concerns the missing of a relevant performance measure. As only a certain part of students in both samples had yet to pass an exam in introductory statistics, sufficiently robust data were not available. Further validation studies should analyze the relation between the WAESTA scores and suitable measures of students' actual statistics performance. Especially, this relation should be most instructive – in as much as relevant studies commonly reported low to moderate correlations (Bandalos et al., 1995; Finney & Schraw, 2003; Macher, Paechter, Papousek, & Ruggeri, 2012; Sesé et al., 2015; Tremblay, Gardner, & Heipel, 2000; Vigil-Colet et al., 2008; Zeidner, 1991). However, these results do not really indicate a general flaw in the measures' criterion validity. Rather, they reflect the motivational consequences of statistics anxiety within a strongly restricted setting (Pekrun, 1988). As the successful passing of statistical requirements in the Master's degree is mandatory, the students' increasingly experienced worry, avoidance tendencies, and feelings of apprehension could dispose them to strenghten their learning effort in order to avoid an impending failure outcome (Macher, Papousek, Ruggeri, & Paechter, 2015; Martin & Marsh, 2003). Accordingly, for the WAESTA scale, also a moderate relation with the students' statistical performance should be assumed.

Finally, as both samples in this study were small and predominantly female, gender was not included in the validation analyses. Relevant findings in the field could consistently demonstrate the females to report a higher level of statistics anxiety (Benson, 1989; Hong & Karstensson, 2002; Macher et al., 2012; Onwuegbuzie & Wilson, 2003). Interestingly, despite the apparently heightened anxiety level of female students, some studies could not substantiate any significant disadvantage in their exam performance (Bradley & Wygant, 1998; Macher, Paechter, Papousek, Ruggeri, Freudenthaler, & Arendasy, 2013). This finding needs further clarification with respect to the underlying motivational and behavioral processes. Hence, female students might have overrated their individually existing anxiety level (Zeidner, 1998) – possibly due to a self-derogatory gender stereotyping effect (Bieg, Goetz, Wolter, & Hall, 2015; Pomerantz, Altermatt, & Saxon, 2002). As well, pursuing a more adaptive coping strategy to avoid feared failure, they might have ramped up their learning approach (Martin & Marsh, 2003). Given a larger sample size with a more adequately balanced gender ratio, this issue should also be examined with respect to the WAESTA scale.

In summary, the present findings yield important information concerning the internal and external validity of the newly developed WAESTA scale. However, they must be seen as preliminary in nature. Therefore, they should represent just a very first step in method development.

References

Ahmed, W., Minnaert, A., Kuyper, H., & van der Werf, G. (2012). Reciprocal relationships between math self-concept and math anxiety. *Learning and Individual Differences, 22,* 385-389. doi: 10.1016/j.lindif.2011.12.004

Baloğlu, M. (2002). Psychometric properties of the Statistics Anxiety Rating Scale. *Psychological Reports, 90,* 315-325. doi: 10.2466/pr0.2002.90.1.315

Bandalos, D. L., Yates, K., & Thorndike-Christ, T. (1995). Effects of math self-concept, perceived self-efficacy, and attributions for failure and success on test anxiety. *Journal of Educational Psychology, 87*, 611-623. doi: 10.1037/0022-0663.87.4.611

Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory model. *Journal of Experimental Education*, *57*, 247-261. doi:10.1080/00220973.1989.10806509

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, *17*, 10-17, 22. doi: 10. 1111/j.1745-3992.1998.tb00616.x

Benson, J., Bandalos, D., & Hutchinson, S. (1994). Modeling test anxiety among men and women. *Anxiety, Stress, and Coping, 7,* 131-148. doi: 10.1080/10615809408249340

Beurze, S. M., Donders, A. R. T., Zielhuis, G. A., de Vegt, F., & Verbeek, A. L. M. (2013). Statistics anxiety: A barrier for education in research methodology for medical students? *Medical Science Educator, 23,* 377-384. doi: 10.1007/BF03341649

Bieg, M., Goetz, T., Wolter, I., & Hall, N. (2015). Gender stereotype endorsement differentially predicts girls‘ and boys‘ trait-state discrepancy in math anxiety. *Frontiers in Psychology, 6,* 1404. doi:10.3389/fpsyg.2015.01404

Birenbaum, M., & Eylath, S. (1994). Who is afraid of statistics? Correlates of statistics anxiety among students of educational sciences. *Educational Research, 36,* 93-98. doi: 10.1080/0013188940360110

Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development, 78,* 246-263. doi:10.1111/j.1467-8624.2007. 00995.x

Blankenstein, K.R., Flett, G. L., & Watson, M. S. (1992). Coping and academic problem-solving ability in test anxiety. *Journal of Clinical Psychology, 48,* 37-46.

Bradley, D. R., & Wygant, C. R. (1998). Male and female differences in anxiety about statistics are not reflected in performance. *Psychological Reports, 82,* 245-246. doi: 10.2466/pr0.1998.82.1.245

Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M., & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin, 139,* 655-701. doi: 10.1037/a0029531

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology, 27,* 270-295. doi: 10.1006/ceps. 2001.1094

Chew, P. K. H., & Dillon, D. (2014). Reliability and validity of the Statistical Anxiety Scale among students in Singapore and Australia. *Journal of Tropical Psychology, 4*(7), 1-7. doi: 10.1017/jtp.2014.7

Chiesi, F., & Primi, C. (2009). Assessing statistics attitudes among college students: Psychometric properties of the Italian version of the Survey of Attitudes toward Statistics (SATS). *Learning and Individual Differences, 19,* 309-313. doi: 10.1016/j. lindif.2008.10.008

Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' statistics achievement. *Statistics Education Research Journal, 9*(1), 6-26.

Chin, E. C. H., Williams, M. W., Taylor, J. E., & Harvey, S. T. (2017). The influence of negative affect on test anxiety and academic performance: An examination of the tripartite model of emotions. *Learning and Individual Differences, 54,* 1-8. doi: 10. 1016/j.lindif.2017.01.002

Cruise, R. J., Cash, R. W., & Bolton, D. L. (1985). Development and validation of an instrument to measure statistical anxiety. Paper presented at the annual meeting of the Statistical Education Section. *Proceedings of the American Statistical Association, Chicago 1985* (pp. 92-97). Washington: American Statistical Association.

Cury, F., Da Fonseca, D., Zahn, I., & Elliot, A. (2008). Implicit theories and IQ test performance: A sequential mediational analysis. *Journal of Experimental Social Psychology, 44,* 783-791. doi: 10.1016/j.jesp.2007.07.003

Dauphinee, T. L., Schau, C., & Stevens, J. J. (1997). Survey of Attitudes Toward Statistics: Factor structure and factorial invariance for women and men. *Structural Equation Modeling, 4,* 129-141. doi: 10.1080/10705519709540066

Deffenbacher, J. L. (1980). Worry and emotionality in test anxiety. In I. G. Sarason (Ed.), *Test anxiety: Theory, research and applications* (pp. 111-128). Hillsdale: Erlbaum.

de Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research, 44,* 147-181. doi: 10.1080/ 00273170902794206

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95,* 256-273. doi: 10.1037/ 0033-295X.95.2.256

Dweck, C. S., & Molden, D. C. (2005). Self-theories. Their impact on competence motivation and acquisition. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 122-140). New York: Guilford Press.

Earp, M. S. (2007). *Development and validation of the Statistics Anxiety Measure.* University of Denver, College of Education: Dissertation.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53,* 109-132. doi: 10.1146/annurev.psych.53.100901.135153

Edmundson, E. W., Koch, W. R., & Silverman, S. (1993). A facet analysis approach to content and construct validity. *Educational and Psychological Measurement, 53,* 351-368. doi: 10.1177/0013164493053002005

Eichhorn, J. (2018). *Die Erfassung von studentischen Valenzkognitionen zur Nützlichkeit statistischer Kompetenzen. Entwicklung, psychometrische Prüfung und vorläufige Validierung eines entsprechenden Befragungsinstruments* [Measuring educational science majors‘ utility value of statistical competence. Scale development, psychometric properties, and preliminary validation results]. Leibniz University Hannover, Faculty of Humanities (Institute of Psychology): unpublished Master's thesis.

Elliot, A. J., & McGregor, H. A. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 76*, 628-644. doi: 10.1037/0022-3514.76.4.628

Emmioğlu, E. (2011). *A structural equation model examining the relationships among mathematics achievement, attitudes toward statistics, and statistics outcomes.* Middle East Technical University Ankara, Graduate School of Social Sciences: Doctoral thesis.

Emmioğlu, E., & Capa-Aydin, Y. (2012). Attitudes and achievement in statistics: A meta-analysis study. *Statistics Education Research Journal, 11*(2), 95-102.

Faber, G. (1995). Die Diagnose von Leistungsangst vor schulischen Recht-schreibsituationen: Neue Ergebnisse zu den psychometrischen Eigen-schaften und zur Validität einer entsprechenden Kurzskala [Measuring elementary school children's spelling-specific anxiety: Further psychometric and validation results]. *Praxis der Kinderpsy-chologie und Kinderpsychiatrie, 44*, 110-119.

Faber, G. (2012a). *Elementary school children's spelling-specific self-beliefs. Longitudinal analyses of their relations to academic achievement, school attitudes, and self-esteem.* New York: Nova Science Publishers.

Faber, G. (2012b). Measuring self-perceptions of oral narrative competencies and anxiety in the EFL context. *Electronic Journal of Research in Educational Psychology, 10*, 1343-1382.

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology, 28*, 161-186. doi: 10.1016/S0361-476X(02) 00015-2

Galassi, J. P., Frierson, H. T., & Sharer, R. (1981). Behavior of high, moderate, and low test anxious students during an actual test situation. *Journal of Consulting and Clinical Psychology, 49*, 51-62.

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary Educational Psychology, 48,* 67-84. doi: 10.1016/j.ced psych.2016.09.003

Goetz, T., Frenzel, A. C., Pekrun, R., Hall, N. C., & Lüdtke, O. (2007). Between- and within-domain relations of students' academic emotions. *Journal of Educational Psychology, 99,* 715-733. doi: 10.1037/0022-0663.99.4.715

Goetz, T., Pekrun, R., Hall, N., & Haag, L. (2006). Academic emotions from a social-cognitive perspective: Antecedents and domain-specificity of students' affect in the context of Latin instruction. *British Journal of Educational Psychology, 76,* 289-308. doi: 10.1348/000709905X42860

Gogol, K., Brunner, M., Preckel, F., Goetz, T., & Martin, R. (2016). Developmental dynamics of general and school-subject-specific components of academic self-concept, academic interest, and academic anxiety. *Frontiers in Psychology, 7,* 356. doi: 10.3389 %2Ffpsyg.2016.00356

González, A., Rodrígez, Y., Faílde, J. M., & Carrera, M. V. (2016). Anxiety in the statistics class: Structural relations with self-concept, intrinsic value, and engagement in two samples of undergraduates. *Learning and Individual Differences, 45*, 214-221. doi: 10.1016/j.lindif.2015.12.019

Graham, J. W. (2012). *Missing data. Analysis and design.* New York: Springer.

Green, J., Martin, A. J., & Marsh, H. W. (2007). Motivation and engagement in English, mathematics and science high school subjects: Towards an understanding of multidimensional domain specificity. *Learning and Individual Differences, 17,* 269-279. doi: 10.1016/j.lindif.2006.12.003

Griffith, J. D., Mathna, B., Sappington, M., Turner, R., Evans, J., Gu, L., Adams, L. T., & Morin, S. (2014). The development and validation of the Statistics Comprehensive Anxiety Response Evaluation. *International Journal of Advances in Psychology, 3*(2), 21-29. doi: 10.14355/ijap.2014.0302.01

Guttman, R., & Greenbaum, C. W. (1998). Facet theory: Its development and current status. *European Psychologist, 3,* 13-36. doi: 10.1027//1016-9040.3.1.13

Hagtvet, K., & Benson, J. (1997). The motive to avoid failure and test anxiety responses: Empirical support for integration of two research traditions. *Anxiety, Stress, and Coping, 10,* 35-57. doi: 10.1080/10615809708249294

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7,* 238-247. doi: 10.1037/1040-3590.7.3.238

Hodapp, V., & Benson, J. (1997). The multidimensionality of test anxiety: A test of different models. *Anxiety, Stress, and Coping, 10,* 219-244. doi: 10.1080/10615809708249302

Hong, E., & Karstensson, L. (2002). Antecedents of state test anxiety. *Contemporary Educational Psychology, 27,* 348-367. doi: 10.1006/ceps.2001.1095

Hong, Y.-Y., Chiu, C.-Y., Dweck, C. S., Lin, D. M.-S., & Wan, W. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology, 77,* 588-599. doi: 10.1037/0022-3514.77.3.588

Hox, J. J. (1997). From theoretical concept to survey question. In L. Lyberg, P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz, & D. Trewin (Eds.), *Survey measurement and process quality* (pp. 47-69). Hoboken: Wiley. doi: 10.1002/9781118490013.ch2

İlhan, M., & Cetin, B. (2013). Mathematics Oriented Implicit Theory of Intelligence Scale: Validity and reliability study. *GESJ: Education Science and Psychology*, *25,* 116-134.

Kieffer, K. M., & Reese, R. J. (2009). Measurement of test and study worry and emotionality in college students. A psychometric evaluation of the Test and Study Attitudes Inventory. *Educational and Psychological Measurement, 69,* 303-321. doi: 10.1177/0013164408323231

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd. ed.). New York: Guilford Press.

Kooken, J., Welsh, M. E., McCoach, D. B., Johnston-Wilder, S., & Lee, C. (2016). Development and validation of the Mathematical Resilience

Scale. *Measurement and Evaluation in Counseling and Development, 49,* 217-242. doi: 10.1177/0748175 615596782

Lalonde, R. N., & Gardner, R. C. (1993). Statistics as a second language? A model for predicting performance in psychology students. *Canadian Journal of Behavioural Science, 25,* 108-125. doi: 10.1037/h0078792

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83,* 1198-1202. doi: 10.1080/01621459.1988.10478722

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4,* 84-99. doi: 10.1037/1082-989X.4.1.84

Macher, D., Paechter, M., Papousek, I., & Ruggeri, K. (2012). Statistics anxiety, trait anxiety, learning behavior, and academic performance. *European Journal of Psychology of Education, 27,* 483-498.

Macher, D., Paechter, M., Papousek, I., Ruggeri, K., Freudenthaler, H. H., & Arendasy, M. (2013). Statistics anxiety, state anxiety during an examination, and academic achievement. *British Journal of Educational Psychology, 83,* 535-549. doi: 10.1111/j.2044-8279.2012.02081.x

Macher, D., Papousek, I., Ruggeri, K., & Paechter, M. (2015). Statistics anxiety and performance: blessings in disguise. *Frontiers in Psychology, 6,* 1116. doi: 10.3389/ fpsyg.2015.01116

Mandler, G., & Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology, 47,* 166-173. doi: 10.1037/h0062855

Marsh, H. W., & O'Mara, A. J. (2008). Self-concept is as multidisciplinary as it is multidimensional. A review of theory, measurement, and practice in self-concept research. In H. W. Marsh, R. G. Craven & D. M. McInerney (Eds.), *Self-processes, learning, and enabling human potential. Dynamic new approaches* (pp. 87-115). Charlotte: Information Age Publishing.

Martin, A. J., & Marsh, H. W. (2003). Fear of failure: Friend or foe? *Australian Psychologist, 38,* 31-38. doi: 10.1080/00050060310001706997

Möller, J., Streblow, L., Pohlmann, B., & Köller, O. (2006). An extension to the internal/ external frame of reference model to two verbal and numerical domains. *European Journal of Psychology of Education, 21,* 467-487.

Nasser, F. M. (2004). Structural model of the effects of cognitive and affective factors on the achievement of Arabic-speaking pre-service teachers in introductory statistics. *Journal of Statistics Education, 12*(1), 1-18.

Nolan, M. M., Beran, T., & Hecker, K. G. (2012). Surveys assessing students' attitudes toward statistics: A systematic review of validity and reliability. *Statistics Education Research Journal, 12*(2), 103-123.

Onwuegbuzie, A. J. (2003). Modeling statistics achievement among graduate students. *Educational and Psychological Measurement, 63,* 1020-1038. doi: 10.1177/0013 164402250989

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment and Evaluation in Higher Education, 29,* 3-19. doi: 10.1080/0260293042000160 384

Onwuegbuzie, A. J., & Wilson, V.A. (2003). Statistics anxiety: Nature, etiology, antecedents, effects and treatments – a comprehensive review of the literature. *Teaching in Higher Education, 8,* 195-209. doi: 10.1080/1356251032000052447

Papanastasiou, E. C., & Zembylas, M. (2008). Anxiety in undergraduate research method courses: its nature and implications. *International Journal of Research and Method in Education, 31,* 155-167. doi: 10.1080/17437270802124616

Papousek, I., Ruggeri, K., Macher, D., Paechter, M., Heene, M., Weiss, E.M., Schulter, G., & Freudenthaler, H. (2012). Psychometric evaluation and experimental validation of the Statistics Anxiety Rating Scale. *Journal of Personality Assessment, 94,* 82-91. doi: 10.1080/00223891.2011.627959

Pekrun, R. (1988). Anxiety and motivation in achievement settings: towards a system-theoretical approach. *International Journal of Educational Research, 12,* 307-323. doi: 10.1016/0883-0355(88)90008-0

Pekrun, R., Elliot, A. J., & Maier, M. A. (2009). Achievement goals and achievement emotions: Testing a model of their joint relations with academic performance. *Journal of Educational Psychology, 101,* 115-135. doi: 10.1037/a0013383

Pekrun, R., Goetz, T., Perry, R. P., Kramer, K., Hochstadt, M., & Molfenter,

S. (2004). Beyond test anxiety: Development and validation of the Test Emotions Questionnaire (TEQ). *Anxiety, Stress, and Coping, 17,* 287-316. doi: 10.1080/10615800412 331303847

Perepiczka, M., Chandler, N., & Becerra, M. (2011). Relationship between graduate students' statistics self-efficacy, statistics anxiety, attitude toward statistics, and social support. *The Professional Counselor, 1*(2), 99-108. doi:10.15241/mpa.1.2.99

Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed. Gender differences in academic performance and internal distress. *Journal of Educational Psychology, 94,* 396-404. doi: 10.1037/0022-0663.94.2.396

Putwain, D. W. (2008). Deconstructing test anxiety. *Emotional and Behavioural Difficulties, 13,* 141-155. doi: 10.1080/13632750802027713

Putwain, D. W., & Symes, W. (2012). Achievment goals as mediators of the relationship between competence beliefs and test anxiety. *British Journal of Educational Psychology, 82,* 207-224. doi: 10.1111/j.2044-8279.2011.02021.x

Rost, D. H., & Schermer, F. J. (1989). The various facets of test anxiety: A subcomponent model of test anxiety measurement. In R. Schwarzer, H. M. van der Ploeg & C. D. Spielberger (Eds.), *Advances in test anxiety research. Volume 6* (pp. 36-52). Lisse: Swets and Zeitlinger.

Rost, D. H., Sparfeldt, J. R., & Schilling, S. R. (2007). *DISK-Gitter mit SKSLF-8. Differentielles Selbstkonzept-Gitter mit Skala zur Erfassung des Selbstkonzepts schulischer Leistungen und Fähigkeiten* [The self-concept grid: A questionnaire for measuring subject-specific facets of academic self-concept]. Göttingen: Hogrefe.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: Reactions to tests. *Journal of Personality and Social Psychology, 46,* 929-938. doi: 10.1037/0022-3514. 46.4.929

Schunk, D. H., & Zimmerman, B. J. (2006). Competence and control beliefs: Distinguishing the means and ends. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology (*2nd ed., pp. 349-367). New York: Erlbaum.

Schutz, P. A., Drogosz, L. M., White, V. E., & Distefano, C. (1998). Prior knowledge, attitude, and strategy use in an introduction to statistics course. *Learning and Individual Differences, 10,* 291-308. doi: 10.1016/S1041-6080(99)80124-1

Schwarzer, R. (1996). Thought control of action: Interfering self-doubts. In I. G. Sarason, G. R. Pierce & B. R. Sarason (Eds.), *Cognitive interference: Theory, methods and findings* (pp. 99-115). Mahwah: Erlbaum.

Schwarzer, R., & Quast, H.-H. (1985). Multidimensionality of the anxiety experience: Evidence for additional components. In H. M. van der Ploeg, R. Schwarzer & C. D. Spielberger (Eds.), *Advances in test anxiety research. Volume 4* (pp. 3-14). Lisse: Swets and Zeitlinger.

Selkirk, L. C., Bouchey, H. A., & Eccles, J. S. (2011). Interaction among domain-specific expectancies, values, and gender: Predictors of test anxiety during early adolescence. *Journal of Early Adolescence,* 31, 361-389. doi: 10.1177/0272431610363156

Sesé, A., Jiménez, R., Montano, J.-J., & Palmer, A. (2015). Can attitudes towards statistics and statistics anxiety explain students' performance? *Revista de Psicodidáctica, 20,* 285-304. doi: 10.1387/RevPsicodidact.13080

Skaalvik, E. M. (1997). Self-enhancing and self-defeating ego orientation: Relations with task and avoidance orientation, achievement, self-perceptions, and anxiety. *Journal of Educational Psychology, 89,* 71-81. doi: 10.1037/0022-0663.89.1.71

Stappert, A. (2017). *Die Erfassung impliziter Fähigkeitstheorien von Studierenden im Umgang mit statistischen Anforderungen. Entwicklung, psychometrische Prüfung und vorläufige Validierung eines entsprechenden Befragungsinstruments* [Measuring educational science majors‘ implicit theories of statistical competence. Scale development, psychometric properties, and preliminary validation results]. Leibniz University Hannover, Faculty of Humanities (Institute of Educational Psychology): unpublished Master's thesis.

Tremblay, P. F., Gardner, R. C., & Heipel, G. (2000). A model of the relationships among measures of affect, aptitude, and performance in introductory statistics. *Canadian Journal of Behavioural Science, 32,* 40-48. doi: 10.1037/h0087099

von der Embse, N., Jester, D., Roy, D., & Post, J. (2018). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of Affective Disorders, 227,* 483-493. doi: 10.1016/j.jad.2017.11.048

Vanhoof, S., Kuppens, S., Sotos, A. E. C., Verschaffel, L., & Onghena, P. (2011). Measuring statistics attitudes: Structure of the Survey of Attitudes Toward Statistics (SATS-36). *Statistics Education Research Journal, 10*(1), 35-51.

Vigil-Colet, A., Lorenzo-Seva, U., & Condon, L. (2008). Development and validation of the Statistical Anxiety Scale. *Psicothema, 20*, 174-180.

Wigfield, A., Hoa, L. W., & Klauda, S. L. (2009). The role of achievement values in the regulation of achievement behaviors. In D. H. Schunk & B. J. Zimmerman (Eds.), *Motivation and self-regulated learning. Theory, research, and applications* (pp. 169-195). New York: Routledge.

Williams, A. (2014). An exploration of preference for numerical information in relation to math self-concept and statistics anxiety in a graduate statistics course. *Journal of Statistics Education, 22,* 1-16. doi: 10.1080/10691898.2014.11889693

Williams, A. S. (2015). Statistics anxiety and worry: The roles of worry beliefs, negative problem orientation, and cognitive avoidance. *Statistics Education Research Journal, 14*(2), 53-75.

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Educational Psychology, 61*, 319-328. doi: 10.1111/j.2044-8279.1991.tb00989.x

Zeidner, M. (1998). *Test anxiety. The state of the art.* New York: Plenum.

Zonnefeld, V. L. (2015). *Mindsets, attitudes, and achievement in undergraduate statistics courses.* University of South Dakota, Graduate Faculty: Dissertation.

Appendix. WAESTA items for measuring education science students' worry (WR), avoidance (AV), and emotionality (EM) cognitions encountering statistical demands (originally worded in German)

01 WR I will hardly be able to meet my degree program's statistics requirements right away.

05 WR It would be difficult for me to discuss statistical content adequately in my papers.

08 WR I have difficulty understanding statistical content in a lecture.

10 WR If I had to comment on statistical data in a course, I would be worried that I would make a fool of myself.

11 WR If I had to give a presentation including statistical findings in a course, I would hope that no one had any follow-up questions.

12 WR I would hardly be able to present a report on statistical research findings adequately.

14 WR Despite careful preparation for a statistics exam, I would worry about not passing it.

16 WR If I took a statistics course, I would be concerned that I would quickly forget everything I had learned.

03 AV If I could, I would rather take two other courses than do one statistics course.

04 AV When presentation topics are being assigned in the course, I would make sure that I receive a topic that doesn't involve statistics.

06 AV When preparing presentations, I would rather omit anything that has to do with statistics.

17 AV In scientific texts, I would skip over statistical tables and diagrams if possible.

02 EM I would be very uncomfortable if I had to work on a statistical problem.

07 EM I would be quite nervous if I were asked to explain a chart from a research report.

09 EM I would have trouble extracting the relevant information from a table of statistical values.

13 EM I would feel very tense if I had to apply a statistical formula.

15 EM The thought of having to explain a statistical problem in a course makes me quite nervous.

Günter Faber, Dr., is a lecturer at Leibniz University Hannover, Institute of Psychology.

Heike Drexler, Dr., is a lecturer at Leibniz University Hannover, Institute of Psychology.

Alexander Stappert, M.A., is a research assistant at the University of Vechta, Department of Social and Education Sciences.

Joana Eichhorn, B.A., is a student in the Master's Programme of Educational Sciences, Leibniz University Hannover.

Contact address: Günter Faber, Institute of Psychology, Leibniz University Hannover, Schloßwender Straße 1, D-30159 Hannover.
Email: faber@psychologie.uni-hannover.de

Instructions for authors, subscriptions and further details:

http://ijep.hipatiapress.com

How Much Guidance Do Students Need? An Intervention Study on Kindergarten Mathematics with Manipulatives

Erin Horan[1], Martha Carr[2]

1) Association of American Colleges and Universities, USA
2) University of Georgia, USA

Date of publication: October 24th, 2018
Edition period: October 2018 – February 2019

To cite this article: Horan, E., & Carr, M. (2018). How Much Guidance Do Students Need? An Intervention Study on Kindergarten Mathematics with Manipulatives. *International Journal of Educational Psychology*, *7*(3), 286-316. doi: 10.17583/ijep.2018.3672

To link this article: http://dx.doi.org/10.17583/ijep.2018.3672

IJEP – International Journal of Educational Psychology, Vol. 7 No. 3 October 2018 pp. 286-316

How Much Guidance Do Students Need? An Intervention Study on Kindergarten Mathematics with Manipulatives

Erin Horan
Association of American Colleges and Universities

Martha Carr
University of Georgia

Abstract

Research has shown that the efficacy of learning with manipulatives (e.g., fingers, blocks, or coins) is affected by multiple variables, including the amount of guidance teachers provide during learning. However, there is no consensus on how much guidance is necessary when learning with manipulatives. The goal of this study was to examine the optimal level of guidance during instruction with manipulatives. The focus was on the timing and level of guidance. The researcher taught students a lesson on counting from one to 10 with pennies and nickel strips. Kindergarten students were taught over five consecutive days in one of four conditions: high guidance, low guidance, high guidance that transitioned to low guidance, and low guidance that transitioned to high guidance. Results showed no difference in learning across the conditions. These results provide valuable information to teachers on the areas of mathematics that do not require the effort of high guidance.

Keywords: mathematics education, elementary school, manipulatives, guidance

2018 Hipatia Press
ISSN: 2014-3591
DOI: 10.17583/ijep.2018.3672

Hipatia Press
www.hipatiapress.com

IJEP – International Journal of Educational Psychology, Vol. 7 No. 3 October 2018 pp. 286-316

¿Cuánta Guía Necesitan los Estudiantes? Una Intervención de Enseñanza de Matemáticas en Estudiantes de Educación Infantil con Manipulativos

Erin Horan
Association of American Colleges and Universities

Martha Carr
University of Georgia

Resumen

Múltiples estudios han demostrado que la eficacia del aprendizaje con medios manipulativos (por ejemplo, dedos, bloques, o monedas) está relacionada con múltiples variables, incluyendo la guía que proveen las maestras y los maestros durante el aprendizaje. Sin embargo, no existe consenso sobre cuánta guía es necesaria durante el aprendizaje con medios manipulativos. La meta de este estudio fue examinar el nivel óptimo de guía necesaria durante el aprendizaje con medios manipulativos. El estudio se enfocó en los momentos y el nivel de orientación. La investigadora enseñó a estudiantes una lección sobre cómo contar del 1 al 10 haciendo uso de monedas de un centavo y tiras de papel con cinco monedas de un centavo dibujadas a un lado y una moneda de cinco centavos al otro. Durante cinco días consecutivos, la lección se impartió a estudiantes de escuela infantil en una de las siguientes cuatro condiciones: nivel de guía alto, nivel de guía bajo, nivel de guía alto con transición a un nivel de guía bajo, y nivel de guía bajo con transición a un nivel de guía alto. Los resultados no demostraron diferencias en aprendizaje entre las cuatro condiciones. Estos resultados proveen información valiosa para maestras y maestros en las áreas de matemáticas que no requieren el esfuerzo de un nivel alto de guía.

Palabras clave: educación en matemáticas, escuela infantil, manipulativos, guía

2018 Hipatia Press
ISSN: 2014-3591
DOI: 10.17583/ijep.2018.3672

Hipatia Press
www.hipatiapress.com

The utility of manipulatives to support learning has been widely accepted and recommended (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). However, investigations by Carbonneau and her colleagues (Carbonneau & Marley, 2015; Carbonneau, Marley, & Selig, 2013) have shown the efficacy of learning with manipulatives is not consistent and depends on many variables related to the instruction including the level of guidance (e.g., high guidance or low guidance) and the student's prior knowledge. There is evidence that at least some guidance during mathematics instruction is necessary for optimal learning but the literature is unclear as to when teachers should provide guidance and when they should allow students to practice alone without teacher help. In the study described below, we examined students' learning with manipulatives (pennies and nickel strips) with varying levels of guidance. We implemented an experiment in which the amount and timing of guidance with manipulatives was tested using four conditions.

Research on Manipulatives and Guidance

Manipulatives refer to any concrete materials, objects, or drawings used during instruction to support students' learning of number and operations. Manipulatives can be simple, such as counting on fingers or unit blocks, or complex, such as using base ten sticks and blocks. In elementary school mathematics classrooms, students learn to count using individual manipulatives to determine "how many" (National Research Council, 2009). Later, students move on to complex manipulatives that represent values of the base-ten system. In elementary school, manipulatives are incorporated into mathematics curricula to aid students' mathematics reasoning and problem solving skills (e.g., Expressions, Investigations, Saxon).

Sowell (1989) conducted a meta-analysis on the effectiveness of using manipulatives during mathematics instruction and found that using manipulatives was better than not using manipulatives. Younger students, especially, benefitted from using manipulatives as they provide concrete objects to students who may not yet be able to think abstractly (DeLoache, 2000; Uttal, O'Doherty, Newland, Hand, & DeLoache, 2009). Carbonneau, Marley, and Selig (2013) followed up on this research and conducted a meta-analysis of 55 studies that explored the efficacy of teaching with complex manipulatives and found that teaching with manipulatives compared to teaching with abstract symbols showed small to medium sized effects on student learning. The research has shown that manipulatives can aid learning, but there are certain variables (i.e., guidance and prior knowledge) that can mitigate their helpfulness.

The term guidance has been used to describe many types of instructional formats. For example, some research has used the term guidance to describe student-teacher interactions that occur during the learning process (e.g., Terwel, van Oers, van Dijk, & van den Eeden, 2009; Mayer, 2004). Other research uses guidance to describe other aspects of instruction, such as guiding students by providing worked examples, formula sheets, or systematically ordering problems to lead students to insightful learning experiences (e.g., Baroody, Purpura, Eiland, & Reid, 2015; Chen, Kalyuga, & Sweller, 2015). Horan (2017) discussed how guidance is used to describe a variety of instructional components as well as the issues that stem from the lack of a clear definition of guidance. For the purposes of this study, guidance is defined as the interaction between a teacher and students, specifically, the quantity and quality of teachers' responsiveness to students' questions and concerns, and teachers' tendency to promote reflection and critical thought with questions and comments. Examples of high quality interaction include a teacher monitoring student response during problem solving and providing assistance as needed, teachers providing feedback and responding to questions from students, students responding verbally to questions from teachers, and teachers creating opportunities for reflection based on students' performance and needs. In contrast, simply providing performance feedback (i.e., correct or incorrect) that is not responsive to students' needs would be considered low guidance.

Overall, prior research shows support for implementing instruction with manipulatives in the mathematics classroom. However, the research on guidance, especially guidance with manipulatives, is less clear. Further, understanding how prior knowledge impacts guidance with manipulatives adds another variable to investigate.

Guidance and Prior Knowledge

Regarding the effectiveness of guidance when using manipulatives, Laski, Jordan, Daoust, and Murray (2015) summarized general findings on young children's learning with mathematics manipulatives and recommended the use of explicit guidance that relates the concrete manipulatives to the abstract numbers they represent. Providing consistent guidance was found to allow students to devote working memory to understanding the content of the mathematics lesson rather than other, extraneous content. In their meta-analysis, Carbonneau, Marley, and Selig (2013) found that high guidance instruction was associated with higher retention and problem solving performance, while low guidance instruction was associated with higher transfer performance when using manipulatives. Carbonneau et al. (2013) also investigated the impact of age on learning and found that students age 3-6 (preoperational age) struggled more when learning with manipulatives compared with students in the concrete operational age group (7-11) or formal operational age group (12 and older). The authors attributed this finding to young students' tendency to struggle with understanding that objects can represent larger mathematical concepts.

While there is clear evidence high guidance is useful for learning, there is some evidence that there are benefits to implementing lessons with low guidance. Therefore, we are interested in understanding the benefits of high versus low guidance, as well as instruction that transitions the level of guidance during learning (e.g., high to low guidance, low to high guidance). We also look to the research on prior knowledge for learning mathematics, to further understand how prior knowledge may determine the usefulness of high or low guidance on learning.

Support for high guidance. Support for high guidance instruction comes from researchers and theorists who argued that without teacher guidance, students left to their own devices will not learn concepts or, worse, learn the wrong concepts (Rogoff, 1990; Cobb, 1995). Social constructivist theorists posit that high guidance during learning with manipulatives is essential because the manipulatives are culturally-specific, external representations that allow children to count before having an internal representation of number. In order to support the eventual development of an internal representation of numbers, students need guidance to be able to recognize what the concrete manipulatives represent (Bruner, 1966; Vygotsky, 1978).

More recent, empirical research supports implementing high guidance during learning, explaining that exploration without the guidance of an instructor can result in students never interacting with the content to be learned (Mayer, 2004). For example, Terwel, van Oers, van Dijk, and van den Eeden (2009) compared the impact of two problem-solving lessons on student learning of percentages and graphs. In the high guidance condition, fifth grade students were taught through the process of guided co-construction; students and teachers created representations of the percentages through teacher-initiated, guided discussions. In the low guidance condition students were provided with ready-made, completed representations and were not engaged in discussion with the teacher. Controlling for pretests scores, children in the high guidance condition performed better on a posttest and transfer test. This provided support for guided, interactive teaching when students are learning problem solving strategies for percentages and graphs.

Fisher, Hirsh-Pasek, Newcombe, and Golinkoff (2013) described guided instruction as a collaborative construction by students and teachers. In their study, Fisher et al. (2013) taught preschool students properties of shapes in three conditions: free play in which student activity was self-directed with no goals for learning (i.e., low guidance), a guided play condition described as discovery learning with the presence of an active teacher participant, and an instruction condition in which the student observed the instructor talking through the material. The authors found that students in the guided play condition showed improved understanding of shapes over the other two conditions, and those improvements were still observed one week later. They found that for understanding properties of shapes, high guidance, even when scripted, was better than instruction that involved the student passively listening to the teacher or playing alone without any guidance.

Carbonneau and Marley (2015) also worked with preschool students to compare the impact of different levels of guidance. The study investigated the impact of guidance on students' conceptual and procedural knowledge on a quantity discrimination task (which side has more) using manipulatives. In their study, the researcher would make two piles of objects and the child would have a crocodile mouth with instructions that the crocodile should eat the bigger number. After making the piles the researcher would ask, "Which one should the crocodile eat?" In one condition, which the authors labeled high guidance, after the child pointed to the pile the crocodile should eat, the researcher would then read the number sentence represented by the piles and crocodile and correct the child if necessary. In the low guidance condition, the researcher prompted the student to read the number sentence. Carbonneau and Marley (2015) found that students who heard the teacher repeat their explanations and were corrected on their errors improved their conceptual and procedural knowledge more than students who only received prompts to recite the number sentence on their own.

Support for low guidance. While the importance of high teacher guidance is evidenced by prior research, others research shows that high guidance does not always lead to improved performance over low guidance. For example, Sengupta-Irving and Enyedy (2014) compared a guided condition, where the teacher led fifth grade students through the problem solving process via interactive discussion to an unguided, open approach, where students completed the problem without any assistance from the teacher. They found no group differences in learning outcomes between conditions on data analysis and probability.

Further, there are situations where not only is high guidance not any better than low guidance, but low guidance is more effective than high guidance. One reason for this it is important to provide learners with time for their own exploration (e.g., Bruner, 1961; Schwartz, 1992). Low guidance instruction gives learners the opportunity to formulate and understand mathematical concepts on their own, which is important for deeper learning of mathematics knowledge (Piaget, 1977; Fuson, 2009). Low guidance can also avoid the effects of overwhelming students' working memory with too many questions or comments from a teacher (Kroesbergen and Van Luit, 2005). It is important to note that pure discovery learning, where students are left with no guidance or instruction and only materials, has not been found to help students learn; instead, researchers advocate for learning that incorporates some outside assistance in the form of feedback on steps the student is taking or outcome feedback on their answers (Alfieri et al., 2011).

Looking to empirical support for low guidance instruction, Kroesbergen and Van Luit (2005) found low guidance was better than high guidance instruction for students with mild intellectual disabilities who were learning multiplication solution procedures. Kroesbergen and Van Luit (2002) likewise found that students in special education classes benefitted more from low guidance than high guidance, however they found that low performing students not identified as having a learning disability benefitted more from high guidance. These findings indicated that students with learning disabilities may have characteristics that differentiate the impact of guidance on learning. While the current study did not implement research with students with disabilities, we chose these examples to highlight the many variables to consider when researching guidance in mathematics instruction, especially given the limited literature on guidance during mathematics instruction with manipulatives. Additional research, which shows support for low guidance instruction depending on students' prior knowledge, is discussed later with the effects of prior knowledge.

Support for transitioning guidance. Another approach to implementing guidance involves starting with high guidance and then transitioning to low guidance as students gain skill and fluency. This format of ordering guidance was studied by Fuchs et al. (2003) who investigated whether initial high guidance instruction followed by exploratory problem solving is superior to exploration followed by guided instruction. Fuchs et al. (2003) found that problem solving improved for students who had high guidance instruction followed by low guidance problem solving with fully worked examples compared to a high guidance, instruction-only condition. However, high guidance instruction followed by low guidance problem solving with partially worked examples, rather than fully worked examples, was not better than high guidance, instruction-only. These findings showed that the optimal level and timing of guidance may depend on multiple variables, such as the age of students, mathematical topic, and structure and content of the instruction or problems. This points to the need for further research on transitioning levels of guidance.

The role of prior knowledge. Cognitive load theory stipulates that students with less prior knowledge need more guidance so as not to exceed their cognitive load. Students with more domain specific knowledge will not need as much guidance because the information is stored in long term memory (Sweller, Ayres, & Kalyuga, 2011). Guidance should be given to support the acquisition of new knowledge, and not to focus on information that has already been learned because this could confuse the students if conflicting information is given (Kalyuga, 2007). This means teachers need to monitor the amount of guidance to give based on students' prior knowledge and experience with a topic.

Fyfe and Rittle-Johnson (2016) investigated the impact of computer feedback on second grade students' learning of equivalency problems. There were three conditions within computer-based problem solving: no-feedback; immediate accuracy feedback after each problem; and summative, accuracy feedback after all 12 problems were solved. Within each of these three conditions students were grouped as having high or low prior knowledge. The impact of feedback differed as a function of prior knowledge. Students with lower prior knowledge, performed better in the feedback conditions than no feedback conditions on solving equivalency problems. For students with higher prior knowledge, all conditions resulted in improvement on solving equivalency problems.

Jitendra et al. (2013) found a different effect of prior knowledge on learning with high and low guidance. They compared a high guidance condition that utilized schema-based instruction to a low guidance, business-as-usual group. The high guidance condition involved a curriculum in which the teacher prompted students to use think-alouds to encourage monitoring and reflection during problem solving. The low guidance condition involved a school-provided, inquiry-based curriculum, in which students worked alone to develop multiple solutions for an ordered set of problems presented on worksheets. Surprisingly, students with higher pretest scores (high prior knowledge) were found to perform significantly better with the high guidance, schema-based curriculum whereas students with lower pretest scores performed better with the low guidance curriculum. Tournaki (2003) compared performance on mathematics addition tasks for second grade students, half of which were general education students and half of which were students with learning disabilities. For students with learning disabilities, significant improvements from pretest to posttest were only found for students in the high guidance instruction group. General education students improved in both the low and high guidance groups. For both the general education students and the students with learning disabilities significant improvements on the transfer task were only found for students in the high guidance condition.

The results of the studies discussed do not paint a clear picture of the role of prior knowledge. Carbonneau, Marley, and Selig (2013) found that high guidance interventions with manipulatives produced better retention than low guidance interventions, but low guidance interventions produced better transfer than high guidance interventions. Another alternative is to include both high and low guidance in the instruction and determine the optimal sequence of guidance (e.g., Darch, Carnine, & Gersten, 1984; Fyfe, Rittle-Johnson, & DeCaro, 2012). Even further, the optimal level or sequence of guidance may also be influenced by students' prior knowledge (e.g., Jitendra et al., 2013; Tournaki, 2003).

Current Study

In the current study we compared student performance on measures of mathematics achievement after one of four five-day treatments that differed in the amount and/or timing of guidance. In the high guidance condition, students were taught with consistent high guidance for all five days. In the low guidance condition, students were taught with low guidance for all five days. In the high to low guidance condition, students were taught with high guidance for the first two days, low guidance for the last two days, with the third day utilized as a transition day where the researcher limited the guidance but did not eliminate it until day 4. In the low to high guidance condition, students were taught with low guidance for the first two days, high guidance for the last two days, with the third day utilized as a transition day where the researcher added some high guidance questions and comments. Our study specifically investigated four questions:

1. How does student performance on measures of mathematics differ based on teacher guidance when using manipulatives?
2. How does teacher guidance impact kindergarten student performance on a transfer task?
3. How does teacher guidance impact kindergarten student performance on a measure of number sense?
4. To what extent does the effect of teacher guidance differ based on kindergarten student prior knowledge (initial skill)?

Carbonneau et al. (2013) found high guidance was optimal for improving student performance on the task being taught. Research has also shown support for low guidance at some point during instruction, but it is not clear if high guidance should be faded out or if it should come after low guidance instruction (e.g., Kroesbergen and Van Luit, 2002, 2005). Therefore, we predicted that one of the transitioning conditions (high to low or low to high) would be best for impacting student performance on the counting task being taught.

Carbonneau et al. (2013) also found that studies that implemented low guidance interventions with manipulatives had higher effect sizes for transfer than the studies that implemented high guidance interventions with manipulatives. On the other hand, students in low guidance instruction group may not learn at all, and may need guidance from the teacher to learn not just the material, but enough to be able to transfer to another task. Prior studies have found lower achieving students need more guidance to understand the content in order to transfer knowledge (Tournaki, 2003). We were interested in understanding transfer effects because we predicted the transitioning conditions would lead to students learning the counting task at hand, but we wondered if the benefits would also transfer to other situations as other studies have investigated. First, we predicted the consistently low guidance condition would not be the optimal condition for transfer because not all students would be able to learn completely on their own without any guidance. We predicted that the low to high and high to low guidance conditions would lead to better transfer because students would have the opportunity to make meaningful connections on their own. This was also our prediction for posttest performance on the Test of Early Numeracy (TEN) as the TEN can be considered far transfer and the same issues and predictions held for the impact of the different conditions on the TEN.

Based on the review by Kalyuga (2007) on the expertise reversal effect, we hypothesized there would be an interaction effect with prior knowledge. Students with low prior knowledge would perform best with consistent high guidance or high to low guidance to learn with manipulatives. If students are not given enough guidance to start with they may learn information incorrectly or may not know where to begin when exploring with manipulatives alone. Students with high prior knowledge may need consistent low guidance or low to high guidance to learn with manipulatives. These students need time to explore alone and already have enough prior knowledge to do this effectively. Starting with high guidance may confuse students with high prior knowledge.

Method

Participants

Consent forms were distributed to kindergarten students at four elementary schools from a southeastern school district. Students at this school district are comprised of 61% white, 17% Hispanic, 13% black, 5% multi-racial, and 4% Asian. One hundred sixty-seven students consented to participate. Of those, one student was absent during the week of the intervention and one student with special needs could not complete the measures for testing so the final sample was 165 (99 males, 66 females).

The sample was comprised of students who were 71.5% white, 13.3% Hispanic, 12.1% black, and 3% Asian. At the start of the study in fall 2015, the average age of students was 5.56 years, SD=0.36. Students came from three different schools. Seven classes participated from the first school, eight classes participated form the second school, and five classes participated from the third school.

It should be noted that several teachers requested to send only students who might benefit from the intervention so as not to have too many students missing class time. As such, some teachers only sent consent forms home with students of their choosing. While this is beneficial for the purposes of understanding how guidance impacts students not performing as well as their peers in mathematics, this does limit the generalizability of our study. Further, as we do not have an accurate count of how many students were originally recruited and asked to participate, we cannot determine the percent of recruited participants who consented to participate.

Design

Students were randomly assigned to one of the four conditions. The lessons were taught by the first author (referred to as researcher) who designed the study. In the academic year prior to the study, the researcher implemented a pilot study with pre-school students to ensure the feasibility of conducting this study with groups of young students. The lessons took place in conference rooms as available, which are typically limited to five to seven seats. Throughout the school day, the researcher pulled students from class in groups of five to seven. Students from different classes would be combined to form groups and the groups could change from day to day. For example, if one class was in the middle of an important lesson, the researcher would go to other classes to pull other students to form the full group. Teaching took place for six to nine minutes per day for five days.

Materials and procedure

All students were assessed at pretest, posttest, and delayed posttest on a counting with manipulatives task, the Test of Early Numeracy (TEN), and a transfer task. All three tests were given at all three time points. The pretest was administered the week before the intervention took place. The posttest was administered the week after the intervention and a delayed posttest was administered two weeks after the intervention. All pretest, posttest, and delayed posttest measures were individually administered in a quiet area. Administration took approximately 10-15 minutes per student, per testing occasion.

Counting task. The counting tasks were designed to assess student ability to count manipulatives. The counting tasks utilized ten boards which are used to teach kindergarten students the order of numbers as part of the Math Expressions curriculum. It should be noted that Math Expressions was not the curriculum used by the school district. The counting task designed for the pretest was different from the counting task designed for the posttests because students had not yet been introduced to the nickel strips at the time of the pretest and the researcher did not want to provide instruction on the nickel strips until the time of the intervention. For the pretest, students were given a board with the numbers one through ten at the top of the board. Below each number was a column for the student to place pennies to show the value of the number (see Figure 1). The researcher asked students, “Place the number of pennies that are written at the top of the column”. The students used pennies to show the numbers given. For the pretest the researcher asked students to place the correct number of pennies under the columns five, eight, three, one, and six. The nickel strips were not used for the pretest. As this task had students fill in five total columns the scores for the pretest counting task could range from 0-5.

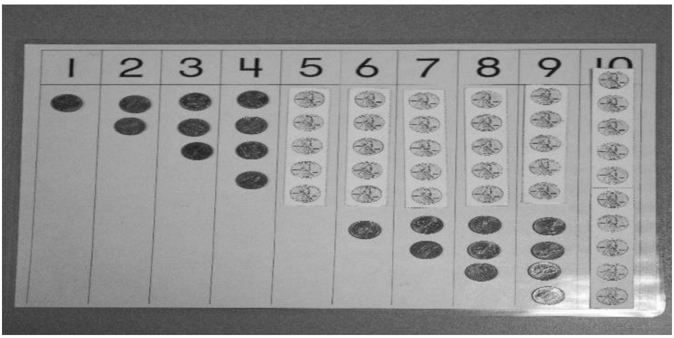

Figure 1. Intervention and the intervention specific task. One through ten board from Math Expressions. The instructions were: "You have a board, pennies and nickel strips. You are going to make the numbers one through ten on your boards using pennies and nickel strips. You can use pennies to make all of the numbers. You can

also use nickel strips. Each nickel strip stands for five pennies (show nickel strip which has pictures of five pennies on it). You can use one nickel strip to take the place of five pennies. The number in the first column is one, can you make the number one with pennies?"

The posttest and delayed posttest had students place all pennies and nickel strips under all columns from one to ten, as shown in Figure 1. This measure was scored as either correct or incorrect which resulted in two categories; 0=not correct; 1=correct. To be scored correctly students needed to place the pennies and nickel strips correctly. The possible range of scores for this task was 0-1. Cronbach's alpha for the pretest, posttest, and delayed posttest for the consistent low guidance condition found counting tasks to be reliable (3 tests; $\alpha = .529$).

Number sense measure. Number sense was measured with the Test of Early Numeracy (TEN). The TEN (Clarke & Shinn, 2004) is individually administered and includes four measures; each measure lasts for one minute for a total of about five minutes per student. The four measures on the TEN are oral counting (possible scores 0-100), number identification (possible scores 0-56), quantity discrimination (possible scores 0-28), and missing number (possible scores 0-21). The oral counting measure has students count as high as they can for one minute. The number identification measure has students identify numbers between 1 and 10 for kindergarteners. The quantity discrimination measure has students identify the larger of two numbers between 1 and 10 for kindergarteners. The missing number measure has students identify the missing number for a set of three numbers with two numbers given. Rather than one, summative score, the TEN yields four separate scores for number sense, which were analyzed individually.

The TEN has been shown to be a valid measure of number sense for kindergarten and first grade students. Clarke and Shinn (2004) found the TEN was correlated with the Woodcock-Johnson Applied Problems (Woodcock & Johnson, 1989) subtest for first grade students, which measures mathematics achievement based on mathematics operations problems and applied mathematics problem. Martinez, Missall, Graney, Aricak, and Clarke (2009) found that the TEN was correlated with Stanford 10 Achievement Test (Harcourt Assessment Inc., 2002), which measures if students are meeting standards for reading, mathematics, and language. Alternate form reliability was measured by testing students with an alternate form of all subtests except for the oral counting measure because there is no alternate form for counting as high as you can (Clarke & Shinn, 2004). Reliabilities for the TEN were measured as .93 for oral counting, .93 for number identification, .92 for quantity discrimination, and .78 for missing number (Clarke & Shinn, 2004). Salvia and Ysseldyke (2001) assigned a reliability of .90 or greater for making educational decisions about individual students, .80 or greater for making screening decisions about individual students, and .60 or greater for making educational decisions about groups of students. According to these guidelines all measures of the TEN can be used to make educational decisions about individuals except the missing number measure, but .78 is still a moderately high reliability.

Transfer task. Transfer was assessed with a task that required students to count on from five. Students were shown a number between six and ten and five circles. Sample transfer problems are shown in Figure 2. Students were given the following instructions: “Do you see that we have 1,2,3,4,5 circles? Can you draw more circles so we have X circles in the box?” This task was scored as zero correct, one correct, or both correct. Cronbach’s alpha for the pretest, posttest, and delayed posttest for the consistent low guidance condition found the transfer task to be reliable (3 tests; α = .667). We chose to report reliability for the consistent low guidance condition as this condition was essentially a control condition; students were not given any guidance during the intervention.

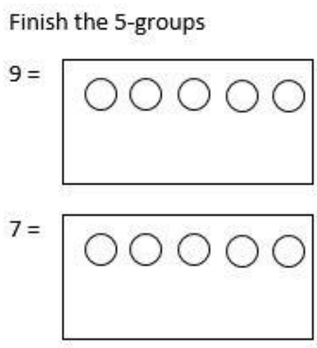

Figure 2. Transfer worksheet for pre-test post-test, and delayed post-test.

Intervention

As described in more detail below, the consistent high guidance group implemented only the high guidance lesson throughout the entire week and the consistent low guidance group implemented only the low guidance lesson throughout the entire week of the intervention. The high to low guidance group began the week with high guidance lessons then shifted to low guidance lessons. The low to high guidance group began the week with low guidance lessons and then shifted to high guidance lessons. To assure fidelity of the high and low guidance modifications, all lessons were recorded and coded as described below.

The four conditions utilized ten boards which are used to teach kindergarten students the order of numbers and are a part of the Math Expressions curriculum (see Figure 1). For this task, students make the numbers one through ten by using pennies and nickel strips. Nickel strips are white pieces of paper that fit perfectly under five pennies. The students first counted out the number of pennies requested then added a nickel strip under sets of five pennies. For example, if the number eight was counted the student would count out eight pennies, then replace five of those pennies with a nickel strip.

High guidance modification. Three of the four conditions include lessons that have high guidance. A high guidance lesson is defined as the teacher asking many questions during learning. A list of possible questions is included in Table 1. The teacher was not required to use every question on this list nor was the list an exhaustive list of questions asked. The high guidance instruction used these questions to increase student learning and understanding. The teacher also provided elaborate feedback about performance during the lesson (not just right or wrong but why), helped students if they needed help, and answered students’ questions.

Table 1
Sample High guidance questions.

"Why did you (not) use a nickel strip in this column?"
"How is this column different from the last column?"
"How is the 8 column the same as the 3 column? How is it different?"
"Can we use a nickel strip in this column? Why (not)?"
"Can you count the pennies to check your answer?"
"How many more pennies would we need for a nickel strip?"
"How many more pennies would we need for another nickel strip?"
"How many more pennies would we need for 5?"
"How many more pennies would we need for 10?"

Low guidance modification. Three of the four conditions include lessons that have low guidance. Per the definition of low guidance for this paper the teacher could provide feedback to students in the form of "yes" or "no" but provided no further information. In the case of the activity to learn the numbers one to ten, low guidance included instructions to make the numbers one to ten and corrective feedback, but did not include any back and forth questioning. In addition to the instructions given in Figure 1, students were provided the following instruction once they reached the five column: "When you reach the number five on the number board you take away the five pennies and use a nickel strip instead." For the numbers six through 10 these instructions were repeated. Questions to keep the students on task could be asked, but questions about the content (e.g., "which number is bigger?") were not.

Transitioning conditions. There were two transitioning conditions; high to low guidance and low to high guidance. For the first two days students were taught with either high or low guidance. Day three was a transition day where the level of guidance started to taper off so that high guidance was tapered to low guidance or increased so that low guidance increased to high guidance. On days four and five students were taught with the second type of guidance so that students who were given low guidance on days one and two were given high guidance and students who were given high guidance on days one and two were given low guidance.

Fidelity. To ensure fidelity of the high and low guidance modification all lessons were audio recorded and coded. Each lesson was rated as high guidance or low guidance based on the number of questions asked by the researcher to students; less than five indicated low guidance, more than five indicated high guidance. Five was chosen as the cutoff to allow room for the low guidance conditions to include minimal questioning such as to keep students on task, as completely cutting out questions is not realistic or practical in everyday teaching. The researcher and a trained independent rater (a graduate student) coded 20% (31) of the sessions. Interrater reliability between the researcher and the independent coder was established as 96.8%.

Results

Performance on Counting with Manipulatives

Means and standard deviations of performance on the pretest, posttest, and delayed posttest counting tasks for students in each condition are shown in Table 2. Analysis of variance showed that pretest scores on the content task did not significantly differ across conditions, F (3,164) = 1.506 p = .225. Posttest scores on the content task controlling for pretest scores were not significantly different across conditions, F (3,160) = 0.735, p = .532. Delayed posttest scores on the content task controlling for pretest scores were not significantly different across conditions, F (3,160) = 1.128, p = .339.

Table 2

Mean Performance on Content Tasks and Standard Deviation.

	H-H	H-L	L-H	L-L	Full sample
Pretest	4.73 (.554)	4.49 (1.00)	4.54 (.745)	4.72 (.701)	4.62 (,769)
Posttest	0.68 (.474)	0.63 (.488)	0.51 (.506)	0.63 (.489)	0.61 (.489)
Delayed Posttest	0.70 (.464)	0.63 (.488)	0.51 (.506)	0.70 (.465)	0.64 (.483)
N	40	41	41	43	165

Note: Pretests were scored as 0-5. Posttests and delayed posttests were scored as 0 (incorrect) or 1 (correct).

Subsequent analyses were performed after removing students who scored with the highest score on the pretest (five out of five) to account for ceiling effects. Means and standard deviations of performance on the pretest, posttest, and delayed posttest counting tasks for students in each condition are shown in Table 3. Again, posttest scores on the content task controlling for pretest scores were not significantly different across conditions, F (3,38) = 0.040, p = .989. Delayed posttest scores on the content task controlling for pretest scores were not significantly different across conditions, F (3,38) = 0.126, p = .944.

Table 3

Performance on Content Tasks with Highest Performers Removed.

	H-H	H-L	L-H	L-L	Full sample
Pretest	3.78 (.441)	3.38 (1.19)	3.64 (.633)	3.29 (.756)	3.53 (.827)
Posttest	0.56 (.527)	0.54 (.519)	0.57 (.514)	0.57 (.535)	0.56 (.502)
Delayed Posttest	0.56 (.527)	0.54 (.519)	0.50 (.519)	0.57 (.535)	0.53 (.505)
N	9	13	14	7	43

Note: Pretests were scored as 0-5. Posttests and delayed posttests were scored as 0 (incorrect) or 1 (correct).

Performance on Transfer

The transfer pretest, posttest, and delayed posttest were two item tasks scored from 0-2. Means and standard deviations of performance on the pretest, posttest, and delayed posttest transfer tasks for students in each condition are shown in Table 4. Analysis of variance showed that pretest scores on the transfer task did not significantly differ across conditions, F (3,164) = 0.283 p = .838. Posttest scores on the transfer task controlling for pretest scores were not significantly different across conditions, F (3,160) = 0.544 p = .653. Delayed posttest scores on the transfer task controlling for pretest scores were not significantly different across conditions, F (3,160) = 0.618 p = .604. Subsequent analyses were performed after removing students who scored with the highest score on the pretest (five out of five) to account for ceiling effects but these findings were not significant.

Table 4

Mean Performance on Transfer Tasks and Standard Deviation Controlling for Pretest.

	H-H	H-L	L-H	L-L	Full sample
Pretest	1.53 (.784)	1.46 (.745)	1.37 (.767)	1.47 (.855)	1.45 (.784)
Posttest	1.63 (.667)	1.61 (.628)	1.44 (.776)	1.63 (.618)	1.58 (.673)
Delayed Posttest	1.68 (.616)	1.59 (.741)	1.71 (.602)	1.60 (.660)	1.64 (.653)
N	40	41	41	43	165

Note: Pretest, posttest, and delayed posttest were scored from 0-2.

Performance on TEN

A three stage hierarchical linear regression was conducted with each posttest TEN score as the dependent variable. Pretest TEN score was entered at stage one of the regression to control for prior knowledge. The four conditions were dummy coded into three variables and were included at stage two. Interactions between the conditions and pretest scores were included at stage three. Performance on each component of the TEN (i.e., counting, missing number, number identification, and quantity discrimination) was analyzed separately.

The hierarchical multiple regression revealed that at stage one, pretest scores on each component of the TEN contributed significantly to the regression model. Beyond stage one, only the model for the counting component of the TEN showed significant contributions by other variables (see Table 5). For the counting component, pretest scores contributed significantly to the regression model, $F\ (1,163) = 326.0$, $p < .001$ and accounted for 66.7% of the variation in posttest scores. Introducing the experimental conditions explained an additional .3% of the variation in posttest scores and this change in R2 was significant, $F\ (4,160) = 81.1$, $p < .001$. Adding the interaction terms to the regression model explained an additional 1.2% of the variation in posttest scores and this change in R2 was significant, $F\ (7,157) = 48.0$, $p < .001$. When all seven variables were included in stage three of the regression model, only two variables were significant predictors of posttest score: pretest score and the interaction between pretest score and the low-high guidance condition. The pretest score uniquely explained 27% of the variation in posttest score and the interaction between pretest and the low-high guidance condition uniquely explained .8% of the variation in posttest score.

Table 5
Summary of Hierarchical Regression Analysis for Variables predicting Posttest Counting

Variable	β	*t*	sr^2	*R*	R^2	ΔR^2
Step 1				0.817	0.667	0.667
Pretest TEN Counting	0.765	18.056*	0.667			
Step 2				0.818	0.670	0.003
Pretest TEN Counting	0.761	17.557*	0.637			
Low-High Condition	0.606	-0.280	0.001			
High-Low Condition	-2.552	-1.178	0.003			
High-High Condition	-1.078	-0.495	0.001			
Step 3				0.826	0.682	0.012
Pretest TEN Counting	0.854	11.597*	0.272			
Low-High Condition	14.541	1.927	0.008			

Table 5 (continued)

High-Low Condition	-0.542	-0.075	0.001
High-High Condition	11.868	1.276	0.003
Low-High*Pretest	-0.240	-2.096*	0.008
High-Low*Pretest	-0.026	-0.236	0.001
High-High*Pretest	-0.192	-1.447	0.004

Note. N=165; * $p < .05$

Figure 3 shows the interaction between counting scores and condition. The graph shows that students with pretest scores below 64.2 (highlighted by the reference line at mean of all pretest scores, Y = 64.2) showed the highest posttest scores in the high to low condition, followed by the consistent high condition, the consistent low condition, and then the high to low condition. Students with pretest scores above 64.2 had the highest posttest scores in the consistent low condition, followed by the high to low condition, the consistent high condition, and finally the high to low condition.

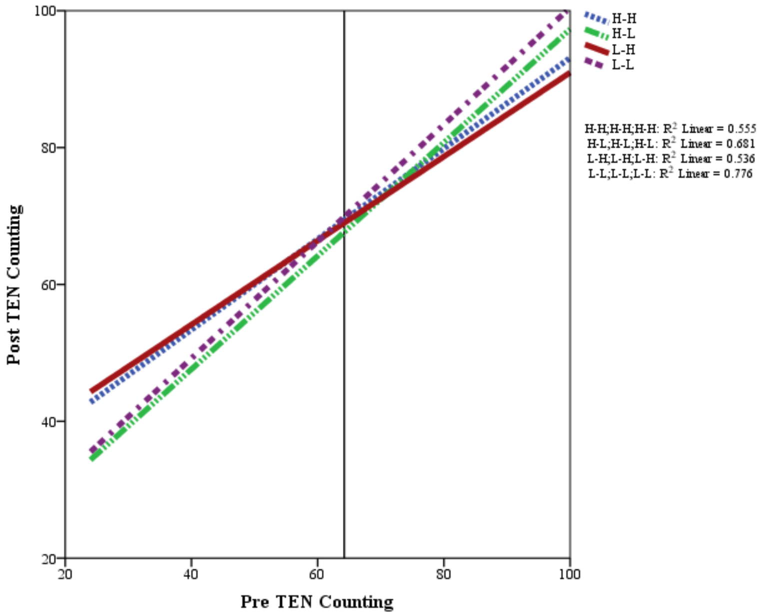

Figure 3. Students' performance on the counting task of the Test of Early Numeracy (TEN) with reference line at mean, *Y* = 64.2.

The same hierarchical linear regression model was performed with delayed TEN counting posttest as the dependent variable but these findings did not remain, only pretest was a significant predictor of delayed posttest score. Subsequent analyses were performed after removing students who scored with the highest score on the counting with manipulatives pretest (five out of five) to account for ceiling effects but these models showed no significant predictors of TEN posttest scores other than TEN pretest scores.

Discussion

Based on the limited prior research on guidance and prior knowledge, we made several hypotheses. We predicted the transitioning high to low and low to high guidance conditions would be the best for students' learning on the intervention content specific task, transfer task, and number sense task. Regarding prior knowledge, we predicted an interaction effect; students with low prior knowledge would perform best on counting with manipulatives with consistent high guidance or high to low guidance on counting with manipulatives. Students with high prior knowledge would perform best with consistent low guidance or low to high guidance when learning with manipulatives. Overall, none of these findings were supported by the results of this study, with most comparisons between conditions showing no difference in learning, even after controlling for prior knowledge.

Overall, student performance on counting with pennies and nickel strips did not differ between conditions even after controlling for pretest scores and possible ceiling effects. These current findings contradict prior research, which has typically shown high guidance groups outperform control groups when highly guided instruction is implemented (e.g., Carbonneau et al., 2013; Hunt, 2014; Terwel et al., 2009). There are several possible explanations for this finding. First, guidance may not be a moderator of learning for counting to ten with manipulatives. This skill may not require explicit explanation or questioning from an instructor. Simply allowing students to practice and count on their own may be all that is needed. Another explanation could be the ineffectiveness of this task for assessing deeper learning. The questions included in the high guidance modifications targeted deeper learning, as they focused on comparing columns and noticing similarities and differences in the quantities. The task assessing the intervention content only had students count pennies and nickel strips, which did not relate to the questions used in the high guidance modifications. Perhaps asking questions to target this deeper learning would have shown differences in learning by condition.

Overall, student performance on the transfer task also did not differ across conditions even after controlling for pretest scores and possible ceiling effects. As with the intervention content specific task, it could be guidance is not a moderator of learning for transfer to counting on from five. However, based on the research on guidance, it is surprising this study did not show differences between conditions. Specifically, the meta-analysis by Carbonneau et al. (2013) suggested that conditions that implemented low guidance (i.e., high to low and low to high) would show greater performance on transfer tasks. Typically, allowing students time to practice while also incorporating guidance (i.e., the high to low or low to high guidance conditions) fosters deeper learning. The contradictory findings of this study could indicate an issue with this measure of transfer as we only included two items to yield a score of 0-2. Perhaps a longer or more in depth test of transfer would have provided better insight into students' learning for transfer.

Student performance on the four tasks for the Test of Early Numeracy showed a difference between conditions for the counting task only, where students counted as high as they could, up to 100, for one minute. After controlling for pretest scores, students in the low to high guidance condition scored significantly lower than students in the consistently low guidance condition. The consistently high and high to low guidance groups did not perform significantly different from the consistently low guidance group. These results indicate that providing students with time to practice alone followed by providing guidance can hinder counting fluency; it was better to allow students to practice counting with manipulative on their own with no additional guidance or questions. These results deviate from prior research that included assessments of number sense, where student performance was significantly higher after high guidance instruction compared to low guidance instruction on counting tasks (e.g., Carbonneau and Marley, 2015, found high guidance positively impacted student's conceptual knowledge as measured by magnitude comparison). Perhaps providing guidance in the form of additional questions can distract students from the main task of counting. But, it is interesting students performed lower in the low to high guidance condition but not the high to low condition.

We hypothesized that after controlling for pretest scores we would find interaction effects; students with higher prior knowledge would excel with less guidance while students with lower prior knowledge would excel with more guidance. As we saw no significant results with any other measures we investigated this hypothesis by graphing scores on the TEN counting task. Figure 3 shows students' pretest and posttest scores on the counting task for each condition. It is clear that students in each condition had a different relationship between pre and posttest performance and this relationship changed dependent upon if students had a pretest score above or below the mean (see reference line Y = 64.2). Looking to students in the low to high guidance condition, we see that for students with lower prior knowledge (i.e., below 64.2) their posttest scores were higher than the other conditions. For students with higher prior knowledge (i.e., above the mean of 64.2) their posttest scores were the lowest of all conditions. This graph brings some clarity to the finding from our hierarchical linear regression that students in the low to high condition performed significantly lower than students in the consistent low guidance condition. Figure 3 shows this clearly for students with higher prior knowledge but not for students with lower prior knowledge. We decided to look at the same hierarchical linear regression but to first analyze students with low prior knowledge (i.e., less than the mean of 64.2) and then high prior knowledge (i.e., greater than the mean of 64.2) but these results revealed no significant predictors of counting scores beyond pretest counting scores.

Based on the overall findings for this study and the tasks used, the only difference in learning was found for the low-high guidance condition, where students performed significantly lower than the consistent low guidance condition on the counting portion of the Test of Early Numeracy. On the TEN counting posttest, students in the low to high guidance condition scored significantly lower when compared to the consistently low guidance condition. However, with this being the only measure to show a difference, we cannot draw conclusions about the optimal level and timing of guidance for learning with mathematics manipulatives.

Conclusions

This research study indicates that level and timing of teacher guidance is not necessary to consider when teaching kindergarten students to count to ten with manipulatives. Controlling for prior knowledge (i.e., pretest) did not impact these results. While not statistically significant, we believe these results are of practical significance to teachers and researchers. These results indicate that the counting tasks already being implemented in the classrooms at this school district are providing enough instruction for this content.

While providing guidance for this task did not appear to impact student learning, these results may not be found for more complex or challenging tasks. The only conclusion we can draw from this study is that for this less complex and age appropriate task, the amount and timing of guidance was not important. Future studies can provide further insight into guidance and its impact on student learning.

Future research

Future research should further advance this research on guidance during instruction as well as mitigate some of the limitations we noted. As discussed earlier, some teachers requested to nominate students who they felt would most benefit from the intervention so as not to pull too many students from class. As such, this reduces our ability to generalize these findings to all students. Another limitation relates to the design of our study. The short duration of the intervention (one week, 6-9 minutes per day) may not be sufficient to impact students' understanding differently between conditions. Further, the measures used in this study were chosen and created to be short, and as such may not have been long or complex enough to determine if deeper learning occurred. Questions that target comparisons between number columns (e.g., "how many more pennies are in the seven column than the five column?") would provide more insight into whether or not deeper, more meaningful learning took place beyond simply counting pennies and nickel strips on a board. We also noted that the lack of statistically significant findings could suggest that the counting tasks already being implemented in the students' classrooms were sufficient for teaching counting from one to ten.

It is also well worth considering the knowledge students are already coming to the classroom with. As such, more in depth pretests could be implemented.

Future research should also focus on variations of the timing and level of guidance with other tasks and age groups. The current research design could also be implemented with preschool students. Preschool students do not have the same base level of knowledge for counting in general and counting with manipulatives specifically. Perhaps implementing this research with preschool students would show differences in learning based on the timing and level of guidance.

Future research could also implement a different research design, such as a single subject design to compare the extent to which guidance is associated with fewer trials to mastery. Given the importance of ensuring fidelity with this research on teacher guidance, single subject design would allow researchers to understand how guidance truly impacts individual students, without the noise created by other students being in the class and impacting student learning. However, given that teaching occurs in classrooms with many students, implementing studies in real classrooms with real teachers is also important to determine the generalizability of these findings.

References

Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tennenbaum, H. R. (2011). Does discovery-based instruction enhance learning? *Journal of Educational Psychology*, *103*(1), 1-18. doi: 10.1037/a0021017

Baroody, A. J., Purpura, D. J., Eiland, M. D., & Reid, E. E. (2015). The impact of highly and minimally guided discovery instruction on promoting the learning of reasoning strategies for basic add-1 and doubles combinations. *Early Childhood Research Quarterly*, *30*(Part A), 93–105. doi: 10.1016/j.ecresq.2014.09.003

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, *31*(1), 21–32.

Bruner, J. S. (1966). *Toward a theory of instruction*. Cambridge, Mass., Belknap Press of Harvard University.

Carbonneau, K. J., Marley, S. C., & Selig, J. P. (2013). A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology*, *105*(2), 380-400. doi: 10.1037/a0031084

Carbonneau, K., & Marley, S. (2015). Instructional guidance and realism of manipulatives influence preschool children's mathematics learning. *The Journal of Experimental Education*, 1–19. https://doi.org/10.1080/00220973.2014.989306

Chen, O., Kalyuga, S., & Sweller, J. (2015). The worked example effect, the generation effect, and element interactivity. *Journal of Educational Psychology*, *107*(3), 689–704. doi: 10.1037/edu0000018

Clarke, B., & Shinn, M. (2004). *Test of early numeracy (ten): Administration and scoring of aimsweb early numeracy measures for use with aimsweb*. Eden Prairie, MN: Edformation Inc.

Cobb, P. (1995). Cultural tools and mathematical learning: A case study. *Journal for Research in Mathematics Education*, *26*(4), 362-385. DOI: 10.2307/749480

Darch, C., Carnine, D., & Gersten, R. (1984). Explicit instruction in mathematics problem solving. *Journal of Educational Research*, *77*(6), 351.

DeLoache, J. S. (2000). Dual representation and young children's use of scale models. *Child Development*, *71*(2), 329–338. doi:10.1111/1467-8624.00148

Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., Golinkoff, R. M. (2013). Taking shape: Supporting preschoolers' acquisition of geometric knowledge through guided play. *Child Development*, *84*(6), 1872-8. https://doi.org/10.1111/cdev.12091

Fuchs, L. S., Fuchs, D., Prentice, K., Burch, M., Hamlett, C. L., Owen, R., … Jancek, D. (2003). Explicitly teaching for transfer: Effects on third-grade students' mathematical problem solving. *Journal of Educational Psychology*, *95*(2), 293–305. https://doi.org/10.1037/0022-0663.95.2.293

Fuson, K. C. (2009). Avoiding misinterpretations of Piaget and Vygotsky: Mathematical teaching without learning, learning without teaching, or helpful learning-path teaching? *Cognitive Development*, *24*(4), 343–361. https://doi.org/10.1016/j.cogdev.2009.09.009

Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology*, *104*(4), 1094-1108. http://psycnet.apa.org/doi/10.1037/a0028389

Fyfe, E. R., & Rittle-Johnson, B. (2016). The benefits of computer-generated feedback for mathematics problem solving. *Journal of Experimental Child Psychology*, *147*, 140–151. https://doi.org/10.1016/j.jecp.2016.03.009

Harcourt Assessment Inc. (2002). *Stanford achievement test, tenth edition*. San Antonio, TX: Harcourt Assessment Inc.

Horan, E. M. (2017). *Guidance and structure in mathematics instruction: How much guidance do students need? An intervention study on kindergarten mathematics with manipulatives* (Doctoral dissertation). Retrieved from https://galileo-usg-uga-primo.hosted.exlibrisgroup.com:443/UGA:UGA:01GALI_USG_ALMA71186079720002931

Hunt, J. H. (2014). Effects of a Supplemental Intervention Focused in Equivalency Concepts for Students with Varying Abilities. *Remedial and Special Education*, *35*(3), 135–144. https://doi.org/10.1177%2F0741932513507780

Jitendra, A. K., Rodriguez, M., Kanive, R., Huang, J.-P., Church, C., Corroy, K. A., & Zaslofsky, A. (2013). Impact of small-group tutoring interventions on the mathematical problem solving and achievement of third-grade students with mathematics difficulties. *Learning Disability Quarterly*, *36*(1), 21-35. doi:10.1177/0731948712457561

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509-539. doi: 10.1007/s10648-007-9054-3

Kroesbergen, E. H., & van Luit, J. E. H. (2002). Teaching multiplication to low math performers: Guided versus structured instruction. *Instructional Science*, *30*(5), 361–378. https://doi.org/10.1023/A:1019880913714

Kroesbergen, E. H., & Van Luit, J. E. H. (2005). Constructivist mathematics education for students with mild mental retardation. *European Journal of Special Needs Education*, *20*(1), 107–116. https://doi.org/10.1080/0885625042000319115

Laski, E. V., Jor'dan, J. R., Daoust, C., & Murray, A. K. (2015). What makes mathematics manipulatives effective? Lessons From cognitive science and montessori education. *SAGE Open*, *5*(2), 1-8. doi: 10.1177/2158244015589588

Martinez, R. S., Missall, K. N., Graney, S. B., Aricak, O. T., & Clarke, B. (2009). Technical adequacy of early numeracy curriculum-based measurement in kindergarten. *Assessment for Effective Intervention*, *34*(2), 116-125. doi: 10.1177/1534508408326204

Mayer, R. E. (2004). Should there be a three-strikes rule against pure discovery learning? *American Psychologist*, *59*(1), 14-19. doi: 10.1037/0003-066X.59.1.14

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for Mathematics*. Washington D.C.: National Governors Association Center for Best Practices, Council of Chief State School Officers. Retrieved from http://www.corestandards.org/the-standards

National Research Council. (2009). *Mathematics Learning in Early Childhood: Paths toward Excellence and Equity* (No. 978-0-3091-2806–3). Washington, D.C.

Piaget, J. (1977). *Epistemology and psychology of functions*. Boston, MA: D. Reidel Publishing Co.

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. New York, NY, US: Oxford University Press.

Salvia, J., & Ysseldyke, J. E. (2001). *Assessment* (*8th ed.*). Boston, MA: Houghton Mifflin.

Schwartz, J. E. (1992). "Silent teacher" and mathematics as reasoning. *Arithmetic Teacher*, *40*(2), 122-124.

Sengupta-Irving, T. & Enyedy, N. (2014). Why engaging in mathematical practices may explain stronger outcomes in affect and engagement: Comparing student-driven with highly guided inquiry. *Journal of the Learning Sciences*, *24*(4), 550-592. https://doi-org.proxy-remote.galib.uga.edu/10.1080/10508406.2014.928214

Sowell, E. J. (1989). Effects of manipulative materials in mathematics instruction. *Journal for Research in Mathematics Education*, *20*(5), 498–505. doi:10.2307/749423

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY : Springer.

Terwel, J., van Oers, B., van Dijk, I., & van den Eeden, P. (2009). Are representations to be provided or generated in primary mathematics education? Effects on transfer. *Educational Research and Evaluation*, *15*(1), 25–44. https://doi.org/10.1080/13803610802481265

Tournaki, N. (2003). The differential effects of teaching addition through strategy instruction versus drill and practice to students with and without learning disabilities. *Journal of Learning Disabilities*, *36*(5), 449-458. https://doi.org/10.1177%2F00222194030360050601

Uttal, D. H., O'Doherty, K., Newland, R., Hand, L. L., & DeLoache, J. (2009). Dual representation and the linking of concrete and symbolic representations. *Child Development Perspectives*, *3*(3), 156–159. doi:10.1111/j.1750-8606.2009.00097.x

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.

Woodcock, R., & Johnson, M. (1989). *Woodcock-Johnson Psycho-Educational Battery-Revised*. DLM Teaching Resources.

Erin Horan, PhD, is a Postdoctoral Research Analyst in the Office of Quality, Curriculum, and Assessment at the Association of American Colleges and Universities (AAC&U).

Martha Carr, PhD, was a University of Georgia (UGA) Aderhold Distinguished Professor of Educational Psychology and Research Fellow of the UGA Institute of Behavioral Research.

Contact Address: Erin Horan, Association of American Colleges and Universities, 1818 R Street NW, Washington DC, USA. Email: ehoran00@gmail.com

Instructions for authors, subscriptions and further details:

http://ijep.hipatiapress.com

Students' Attitude and Motivation in Bilingual Education

Beatriz Calderón Jurado[1] & Cristina Morilla García[1]

1) University of Cordoba, Spain

Date of publication: October 24th, 2018
Edition period: October 2018 – February 2019

To cite this article: Calderón Jurado, B., & Morilla García, C. (2018). Students' attitude and motivation in bilingual education *International Journal of Educational Psychology*, *7*(3), 317-342. doi: 10.17583/ijep.2018.3558

To link this article: http://dx.doi.org/10.17583/ijep.2018.3558

IJEP – International Journal of Educational Psychology, Vol. 7 No. 3 October 2018 pp. 317-342

Students' Attitude and Motivation in Bilingual Education

Beatriz Calderón Jurado, Cristina Morilla García
University of Cordoba

Abstract

Motivation and attitude are two affective factors that can determine the success of students' learning. The objective of this study is to investigate how attitude and motivation affect students in bilingual teaching programs. This study comprised a total of 159 Spanish students and was conducted in the third and fourth year of primary education in which the level of the L2 corresponds to A2.1 of the CEFR (Common European Framework of Reference). The study involved two schools in the province of Cordoba (Andalusia) in which CLIL (Content and Language Integrated Learning) program is implemented. A quantitative methodology has been used whereby students´ attitude and motivation have been analysed through a questionnaire. The results show that although motivation and attitude are positive in science subjects within the bilingual program, it is necessary to pay greater attention to diversity and to the different pace of learning among students.

Keywords: attitude; motivation; bilingual program; CLIL; science subjects.

2018 Hipatia Press
ISSN: 2014-3591
DOI: 10.17583/ijep.2018.3558

Hipatia Press
www.hipatiapress.com

IJEP – International Journal of Educational Psychology, Vol. 7 No. 3 October 2018 pp. 317-342

Actitud y Motivación de los Estudiantes en la Educación Bilingüe

Beatriz Calderón Jurado, Cristina Morilla García
University of Cordoba

Resumen

La motivación y la actitud son dos factores afectivos que pueden determinar el éxito del aprendizaje de los alumnos. El objetivo de este estudio es indagar cómo la actitud y la motivación afectan a los estudiantes con respecto al programa bilingüe. Esta investigación comprende un total de 159 estudiantes españoles y ha tenido lugar en el tercer y cuarto año de educación primaria en el que el nivel de la L2 corresponde al A2.1 del MCER (Marco Común Europeo de Referencia). El estudio se ha realizado en dos centros de la provincia de Córdoba (Andalucía) en los que el programa AICLE (Aprendizaje Integrado de Contenidos e Idiomas) está implementado. Se ha aplicado una metodología cuantitativa en la que se ha analizado la actitud y la motivación de los estudiantes mediante un cuestionario. Los resultados muestran que aunque la motivación y la actitud son positivas en las asignaturas de ciencias dentro del programa bilingüe, es necesario reforzar la atención a la diversidad y los diferentes ritmos de aprendizaje.

Palabras clave: actitud; motivación; programa bilingüe; CLIL; asignaturas de ciencias.

2018 Hipatia Press
ISSN: 2014-3591
DOI: 10.17583/ijep.2018.3558

Nowadays, English is a language that is spoken around the world, according to Broughton, Brumfit, Pincas and Wilde (2002), there are 300 million native speakers in every continent of the globe. In addition, there are over 250 million speakers who use English as a second language in their daily life: it is the language of mass media, official institutions and the language of the commercial and industrial organizations. Consequently, it has become a necessity to learn English at school. The target language has been integrated into the Spanish education system in such a way that educational programs have become bilingual in nature. These programmes require students to become familiar with the foreign language while learning subjects such as science, art and physical education among others. Therefore, affective factors such as attitude and motivation can play an important role in the process of acquiring the second language. According to Latchanna and Dagnew (2009), attitude is seen as a relevant concept for understanding human behaviour and is defined as a mental state which includes feelings and beliefs. Beliefs are connected with success in language lessons. Following Lennartsson (2008), motivation and the desire to learn a second language are considered more relevant factors than social ones; positive attitudes among students tend to raise students´ motivation. In fact, a good teacher should tap into the sources of intrinsic motivation and try to find ways to relate them to external motivation factors. Likewise, it is appropriate to identify learners´ objectives and needs in order to develop accurate motivational strategies. Appropriate attitudes and feelings are needed to raise the competence of students in language learning lessons. The aim of this research is to investigate the value of affective factors such as attitude and motivation in bilingual education.

The competence of attitude and motivation in bilingual education

Learning a foreign language is a complex process in which several factors are involved. The process of learning a foreign language not only implies a cognitive approach. Rather, there are several factors which influence the learning of a foreign language, such as the context, learning achievement, intelligence, age, etc. and affective factors as attitude, motivation and anxiety. Focusing on attitude, Gardner and Lambert (1972) demonstrated that the ability of students to achieve proficiency in a second language is not only influenced by mental competence or language skills, but also on the learners´ attitudes and perceptions about the target language. Furthermore, Victori and Lockhart (1995) asserted that students´ negative beliefs are related to class anxiety, low cognitive achievement and negative attitudes. Following this line, Reid (2003) stated that attitude is related to the achievement of the student due to the fact that learning a foreign language implies not only the intellectual capacity of the learner but also the attitude towards language learning. In addition, Montano and Kasprzyk (2008) assert that:

> Attitude is determined by the individual´s belief about outcomes or attributes of performing the behaviour (behavioural beliefs), weighted by evaluations of these outcomes or attributes. Thus, a person who holds strong beliefs that positively valued outcomes will result from performing the behaviour. Conversely, a person who holds strong beliefs that negatively valued outcomes will result from the behaviour will have a negative attitude (p.71).

The same happens with motivation. In fact, there are several models based on motivation and language learning. Gardner and Lambert (1972) established the socio-psychological model: for them, there are two kinds of motivation: the first kind is integrative motivation which refers to learners´ willingness to learn the language in order to take part in the community which speaks that language. The second kind is instrumental motivation, which deals with learners´ desire to acquire foreign language proficiency for practical purposes. Another framework concerns Self-Determination Theory (Deci & Ryan, 1985; Noels, 2001; Noels, Pelletier, Clément, & Vallerand, 2003). This theory covers also two kinds of motivation: extrinsic motivation, which deals with external factors that influence the learning of a foreign language; and intrinsic motivation, which refers to the interest of learners generated by doing an activity. According to Dörnyei (1998), motivation affects success rate of learners. In fact, if motivation is lacking, even the most remarkable learners with the most impressive abilities are unable to achieve a long-term goal. Likewise, motivation plays an important role in the acquisition of a foreign language as it is a communication coding taught at school. It is an integral part of individual identities, as well as being a form of social organisation that is embedded in a culture community.

Content Language and Integrated Learning (CLIL) in Andalusia

According to García (2009) teachers have a responsibility to educate students as responsible citizens who are prepared for the globalized world. For these reasons it is essential to implement bilingual education in Spain and to follow a suitable approach for the acquisition of the lingua franca. According to Peter Mehisto (2012), and based on the definition provided by Maljers, Marsh and Wolff (2007), CLIL is a dual-focused teaching and learning approach in which the L1 and the additional language or two are used for promoting both content mastery and language acquisition to pre-defined levels. Several subjects can be taught through the CLIL program depending on the qualification of teachers. Nowadays the focus is on history, geography, science and social sciences specifically in Secondary Education. According to Pérez-Cañado (2012), the materials used for these subjects can be adapted from authentic sources or designed with the assistance of Information and Communication Technology (ICT). Coyle (2010) established the 4C Framework, which are, according to him, the four basic ingredients for the CLIL classroom: content, communication, culture and cognition.

Method

Objectives and Research Questions

The aim of this research is to investigate specific research questions concerning the attitude and motivation of students towards their school's bilingual program, more specifically, with regards to science subjects. To do so, a quasi-experimental design has been developed using a questionnaire. According to Campbell and Stanley (1963) a quasi-experimental design is made from situations which already exist in the real world and are more representative in an educational context. It is difficult to control many of the variables and the groups cannot be modified. Specific objectives, attitude and motivation have been evaluated in relation to the following research questions:

1. Are students feeling motivated with the bilingual program at school?
2. Is this bilingual program improving their attitude towards English?

3. Do children value the fact that they can learn English in subjects such as science?
4. How do they feel when the teacher makes use of the second language in the classroom?

Contextualization and Data Gathering

The present study was carried out in two different public education schools of primary education. The first one is the CEIP 14001682, which is situated in center Córdoba. The second one is the CEIP 14004075, which is situated in a town in northern Córdoba. Number of students and schedules for bilingual classes in every school are presented in table 1 and table 2.

Table 1
Number of students and schedule CEIP 14001682

Groups	Schedule	Number of students
3rd A	1rst TIME	26
4th A	2nd TIME	24
4th B	3rd TIME	22
4th D	4th TIME	24
3th B	5th TIME	26

Table 2
Number of students and schedule CEIP 14004075

Groups	Schedule	Number of students
3rd A	4th TIME	21
4th A	5th TIME	16

The instrument employed for data collection was a questionnaire, named “Attitude and motivation in bilingual education within the primary school”.

This instrument was revised for three experts specialized in the subject of the area in order to validate it. The Delphi method was used: a) one EFL professor, b) one applied linguistics university lecturer, and c) one computer engineer specialized in e- learning. The questionnaire is composed of 21 questions divided into four dimensions: students´ motivation towards the bilingual program, students´ attitude towards the bilingual program, assessment of students about learning English in subjects such as sciences, and the use of the second language in the classroom. Each item is rated on a 5-point Likert scale ranging from 1 (strongly agree) to 5 (strongly disagree). The data obtained is quantitative in nature; however, as responses were descriptive, qualitative data is also offered. Besides, Alpha´s Cronbach has been used, which demonstrates the internal reliability of the questionnaire (See table 3). This statistical method gives us information about "the degree to which the items in a scale measure similar characteristics" (Pérez-Paredes & Martínez-Sanchez, 2000-2001, p. 341). The more the results are approximated to .1, the more reliable the questionnaire; the developed questionnaire yields an internal consistency of .726, which means that the reliability of the questionnaire is acceptable.

Table 3
Reliability statistics

Alpha´s Cronbach	Alpha´s Cronbach based on standardized items	N of items
,691	,726	23

Results

The results and data analyses presented below are grouped into four dimensions in order to answer the research questions respectively.

The first dimension addresses students´ motivation towards the bilingual program, and the obtained results respond to the first research question.

Table 4
Valid percentage questionnaires´ results, question 1
"Students are pleased when using the foreign language among themselves".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	116	73,0	73,0	73,0
	I agree	21	13,2	13,2	86,2
	I do not know	17	10,7	10,7	96,9
	I disagree	3	1,9	1,9	98,7
	I disagree completely	2	1,3	1,3	100,0
	Total	159	100,0	100,0	

As it can be shown in table 4, children demonstrate a positive attitude towards speaking English among themselves; additionally, in table 5, it is observed that the use of the ICT in lessons makes students appreciate learning the lingua franca. Nowadays ICTs are a great tool for teaching English. There are many applications and games that help students to learn English and improve their attitude and motivation towards the second language.

Table 5
Valid percentage questionnaires´ results, question 2
"Students value the use of ICTs for learning contents".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	115	72,3	72,3	72,3
	I agree	32	20,1	20,1	92,5
	I do not know	8	5,0	5,0	97,5
	I disagree	1	,6	,6	98,1
	I disagree completely	3	1,9	1,9	100,0
	Total	159	100,0	100,0	

Table 6

Valid percentage questionnaires´ results, question 3

"The use of the mother tongue in lessons makes students feel confident when they are unable to understand the concept in English".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	101	63,5	63,5	63,5
	I agree	15	9,4	9,4	73,0
	I do not know	8	5,0	5,0	78,0
	I disagree	14	8,8	8,8	86,8
	I disagree completely	21	13,2	13,2	100,0
	Total	159	100,0	100,0	

Furthermore, in table 6 it is exposed that more than half of students (63,5%) completely agree, saying that they feel more confident when the teacher uses the mother tongue to clarify something they have not understood in English. It may be that sometimes students are unable to properly understand vocabulary in English and that they need further explanations. In addition to that, as it is analyzed in table 7, 61,6 % of students completely agree considering English useful as a valuable tool for the future, and, this idea has a positive effect on their attitude and motivation.

Table 7

Valid percentage questionnaires´ results, question 4

"Students approve learning science subjects within the bilingual program".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	98	61,6	61,6	61,6
	I agree	21	13,2	13,2	74,8
	I do not know	18	11,3	11,3	86,2
	I disagree	6	3,8	3,8	89,9
	I disagree completely	16	10,1	10,1	100,0
	Total	159	100,0	100,0	

Table 8

Valid percentage questionnaires´ results, question 5

"Students value the possibility of practising English at home with their parents".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	113	71,1	71,1	71,1
	I agree	17	10,7	10,7	81,8
	I do not know	16	10,1	10,1	91,8
	I disagree	6	3,8	3,8	95,6
	I disagree completely	7	4,4	4,4	100,0
	Total	159	100,0	100,0	

Finally, it needs to be mentioned that table 8 is representative by showing that 71,1% of learners agree completely and appreciate their parents´ help with their English homework. This constitutes great scaffolding by parents and increases children's motivation to continue learning English.

The second dimension that is analyzed is the students´ *attitude* towards the bilingual program that addresses the second research question of the study.

Table 9

Valid percentage questionnaires´ results, question 6

"Students feel satisfaction when the teacher uses the lingua franca".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	86	54,1	54,1	54,1
	I agree	43	27,0	27,0	81,1
	I do not know	21	13,2	13,2	94,3
	I disagree	7	4,4	4,4	98,7
	I disagree completely	2	1,3	1,3	100,0
	Total	159	100,0	100,0	

It is appreciated in table 9 that most of the students demonstrate a positive attitude towards the use of the second language when the teacher is explaining the lesson. This is a positive result because English is a language of communication and within the bilingual the program is necessary to teach at least 50-70% of the lessons in English.

Table 10

Valid percentage questionnaires´ results, question 7

“Learning science subjects through English is beneficial for us”.

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	84	52,8	52,8	52,8
	I agree	41	25,8	25,8	78,6
	I do not know	16	10,1	10,1	88,7
	I disagree	7	4,4	4,4	93,1
	I disagree completely	11	6,9	6,9	100,0
	Total	159	100,0	100,0	

Furthermore, in table 10, the item of "learning science subjects through English is beneficial for us", most of students agree completely (52'8%) or agree (25´8%) with that statement. This is a positive result concerning attitude and motivation as well as the bilingual program. They value the opportunity of learning science subjects through English. The results concerning the ability of students to learn in English (table 11) show that most of them (36,5% and 22,6%) agree. This aspect is related to attitude because it shows that the students enjoy the lessons and, consequently, their attitude is positive.

Table 11

Valid percentage questionnaires´ results, question 8

"Bilingual lessons are accessible to everybody".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	58	36,5	36,5	36,5
	I agree	36	22,6	22,6	59,1
	I do not know	28	17,6	17,6	76,7
	I disagree	12	7,5	7,5	84,3
	I disagree completely	25	15,7	15,7	100,0
	Total	159	100,0	100,0	

The third dimension is related to the third research question of the study, dealing with the assessment of students about learning English in subjects such as sciences.

Results in table 12 indicate that, although learners feel shy when they have to speak in front of the class not using their mother tongue, they enjoy participating in lessons (53,3%). That is, despite feeling embarrassed using it, they are able to use the lingua franca, which is desirable because they are communicating in English. Tables 13 and 14 show that 60,4% and 47,7% of students are highly motivated about learning English. This is a significant result, Gardner (1985, p.10) described that if students show desire when doing an activity, they will have a satisfactory experience that is relevant for language learning.

Table 12

Valid percentage questionnaires´ results, question 9

"Students are pleased with using the second language among peers and the teacher".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	85	53,5	53,5	53,5
	I agree	27	17,0	17,0	70,4
	I do not know	29	18,2	18,2	88,7
	I disagree	6	3,8	3,8	92,5
	I disagree completely	12	7,5	7,5	100,0
	Total	159	100,0	100,0	

Table 13

Valid percentage questionnaires´ results, question 10

"Students really appreciate bilingual lessons"

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	96	60,4	60,4	60,4
	I agree	36	22,6	22,6	83,0
	I do not know	18	11,3	11,3	94,3
	I disagree	5	3,1	3,1	97,5
	I disagree completely	4	2,5	2,5	100,0
	Total	159	100,0	100,0	

Table 14
Valid percentage questionnaires´ results, question 11
"Students wait enthusiastically for bilingual lessons"

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	76	47,8	47,8	47,8
	I agree	30	18,9	18,9	66,7
	I do not know	30	18,9	18,9	85,5
	I disagree	7	4,4	4,4	89,9
	I disagree completely	16	10,1	10,1	100,0
	Total	159	100,0	100,0	

Table 15
Valid percentage questionnaires´ results, question 12
"Learners would like to have more subjects within the bilingual program".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	52	32,7	32,7	32,7
	I agree	25	15,7	15,7	48,4
	I do not know	17	10,7	10,7	59,1
	I disagree	18	11,3	11,3	70,4
	I disagree completely	47	29,6	29,6	100,0
	Total	159	100,0	100,0	

Moreover, as we can appreciate in table 15, 32,7% of the students disagree completely with this statement. This can be produced by two factors. On the one hand, we have to take into account the level of difficulty of the subjects, it might be more complicated to learn concepts in the foreign language. The second factor is related to the quantity of subjects offered inside the bilingual program. For instance, CEIP 14001682 includes physical education, music and science in its bilingual program; results in table 16 indicate that 61,6% of students completely agree; they consider English useful as a valuable tool for the future, and, this idea has a positive effect on their attitude and motivation.

Table 16

Valid percentage questionnaires´ results, question 13

"Students approve learning science subjects within the bilingual program"

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	98	61,6	61,6	61,6
	I agree	21	13,2	13,2	74,8
	I do not know	18	11,3	11,3	86,2
	I disagree	6	3,8	3,8	89,9
	I disagree completely	16	10,1	10,1	100,0
	Total	159	100,0	100,0	

Lastly, concerning the role of the families, the analysis in table 17 shows that 68,8% of the children agrees completely. This implies motivation on the part of both parents and students and that they consider English a useful tool for the future.

Table 17

Valid percentage questionnaires´ results, question 14

"Students´ families value the opportunity to learn English through the bilingual program".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	109	68,6	68,6	68,6
	I agree	18	11,3	11,3	79,9
	I do not know	25	15,7	15,7	95,6
	I disagree	2	1,3	1,3	96,9
	I disagree completely	5	3,1	3,1	100,0
	Total	159	100,0	100,0	

The last dimension that is analyzed concerns "the use of the second language in the classroom". The results obtained respond to the fourth research question of our study. As can be observed in table 18, understanding the teacher using the second language, generates varied results. Some students (25,5%) agree completely with that affirmation, whereas others (27%), disagree completely. However, it is advisable to help students to feel comfortable with the second language and make use of it. Likewise, the obtained results in table 19 show that 85,5% of the students feel grateful for having a teaching assistant at school: they can help students to promote English as a medium of communication.

Table 18

Valid percentage questionnaires´ results, question 15

"Students feel weak when they are unable to understand contents which are explained in the second language".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	40	25,2	25,2	25,2
	I agree	26	16,4	16,4	41,5
	I do not know	27	17,0	17,0	58,5
	I disagree	23	14,5	14,5	73,0
	I disagree completely	43	27,0	27,0	100,0
	Total	159	100,0	100,0	

Table 19

Valid percentage questionnaires´ results, question 16

"Students appreciate the role of the teaching assistant in the lessons".

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	128	80,5	80,5	80,5
	I agree	14	8,8	8,8	89,3
	I do not know	11	6,9	6,9	96,2
	I disagree	1	,6	,6	96,9
	I disagree completely	5	3,1	3,1	100,0
	Total	159	100,0	100,0	

Aside from this, assistant teachers can introduce learners to their culture and by doing so students can gain knowledge about some typical customs of the country of the second language. Finally, results in table 20 shows that, although students like learning science through the L2, they feel shy when it comes to taking part in the bilingual lesson. That is a negative aspect of the bilingual program that needs to be tackled, perhaps by building greater cooperation between teachers and pupils.

Table 20

Valid percentage questionnaires´ results, question 17

“Students feel shy participating on bilingual lessons”.

		Frequency	Percentage	Valid percentage	Accumulated percentage
Valid	I agree completely	43	27,0	27,0	27,0
	I agree	19	11,9	11,9	39,0
	I do not know	17	10,7	10,7	49,7
	I disagree	17	10,7	10,7	60,4
	I disagree completely	63	39,6	39,6	100,0
	Total	159	100,0	100,0	

As a global result, it can be contended that the bilingual program improves students´ attitude and motivation. Learners feel satisfaction when they are learning sciences through English because they know they own the possibility to improve their English quality apart from being English an opportunity for their future career. However, sometimes, when the teacher makes use of the second language in the classroom, they hesitate about their answers. It is relevant to make students feel restored and help them when using the second language.

Furthermore, a comparison between both schools has been carried out taking into account students from the 3rd and 4th years. The following figures show the results of this comparison:

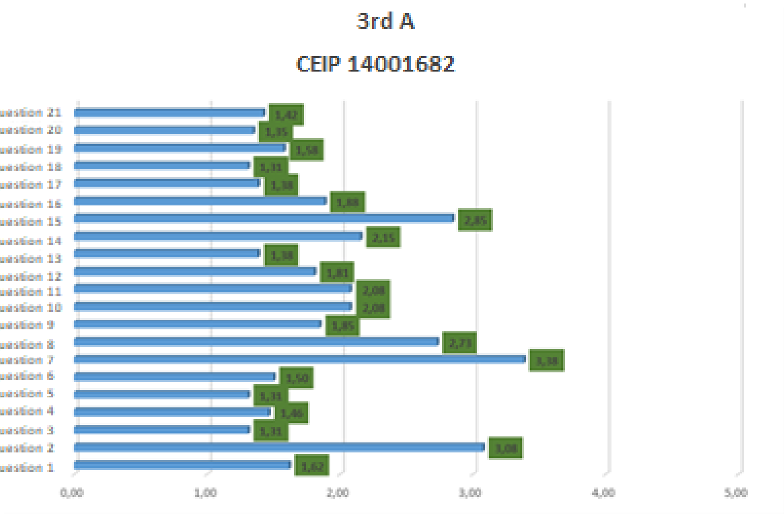

Figure 1. Average questionnaires´ results 3rd A CEIP 14001682

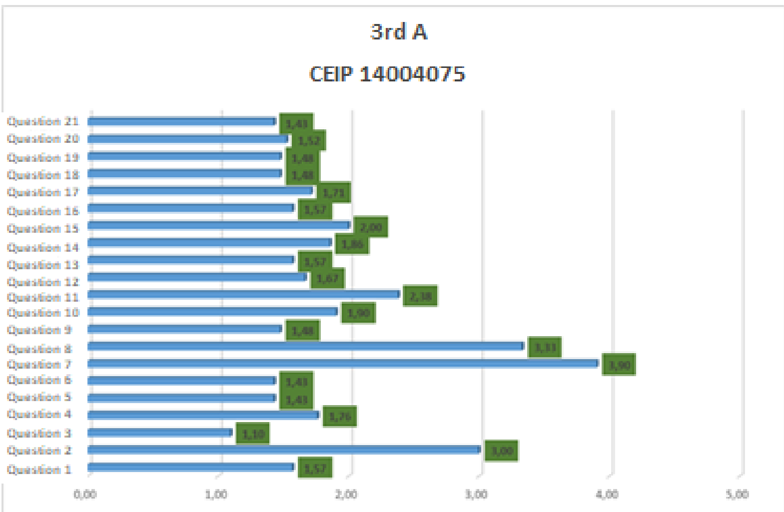

Figure 2. Average questionnaires´ results 3rd A CEIP 14004075

According to the data gathered, the figures above represent the average answers for each question from the questionnaire. Firstly, in 3rd year students from the CEIP 14001682 (see figure 1), the majority of answers were positive, with the most chosen answers being “I completely agree” and “I agree”, as can be observed in the diagram. Regarding 3rd year students from the CEIP 14004075 (see figure 2), most of the results are also positive, with the answers “I completely agree” and “I agree” appearing in response to most of the questions. However, with question number 7 “I do not like when the teacher speaks English in the classroom”, children hesitate about the answer in both institutions, and most of them choose number 3 “I do not know”. Then, in question number 8, “I feel shy when I have to take part in the English lesson”, students from CEIP 14001682 nearly agree whereas students from CEIP 14004075 do not know. Finally, question number 15, related to intrinsic motivation, “I would like to have more subjects in which English is spoken”, learners from CEIP 14001682, are close to “I do not know”, while children from CEIP 14004075 agree. This may be due to the quantity of subjects offered in each bilingual school.

These results, which are reflected in figures 1 and 2 respectively, are positive. Nevertheless, students have some problems when the teacher uses the L2 in the classroom and when they need to use it. Moreover, it is important that students feel comfortable in the lessons, as well as using the second language. In this case, the role of emotional intelligence is relevant. According to Pavón and Ávila (2009, p. 91), emotional intelligence is essential for the acquisition of the L2: it helps to develop the student’s own abilities and enables them to interact within the classroom.

When focusing on students from 4th year (see figure 3 and figure 4), in question number 2, “If I do not understand something that my teacher has explained in English I feel weak”, nearly all students from CEIP 14001682 answered “do not know”, and, students from CEIP 14004075 are close to disagree. If we focus on question number 4, in both institutions there is a positive result towards learning English in subjects such as science. However, in question number 7 “I do not like when the teacher speaks English in the lesson”, students from both schools hesitate. Nevertheless, the 4th year group from CEIP 14001682, seems to be more confident when the teacher speaks the second language in the classroom. Question 12 “I really like to take part in lessons in which English is spoken because I feel comfortable with my peers and my teacher”, students from CEIP 14001682 agree, although students from CEIP 14004075 “do not know”. This difference may be a product of the number of years spent within the bilingual program in each school. Following with question number 14, in both schools, children wait for lessons enthusiastically, meaning that they exhibit positive motivation.

Finally, in question number 15 “I would like to have more subjects in which the English language is spoken” in both primary schools the result is close to “I do not know”, however in the CEIP 14004075, children seem more in favour of having more subjects; currently, they only study science within the bilingual program. According to this analysis, attitude and motivation are favourable among the two groups. It has been demonstrated that attitude and motivation bear an influence on the acquisition of a foreign language. What’s more, if these two affective factors are high, students’ English will improve because attitude and motivation are related to how students feel and behave.

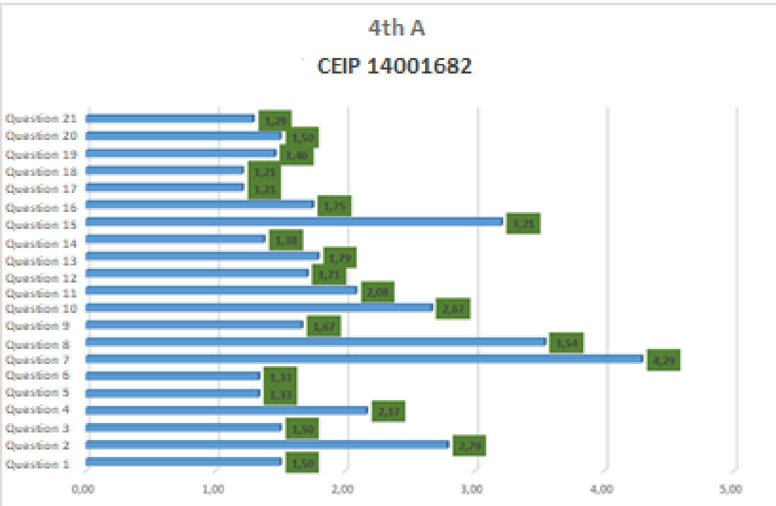

Figure 3. Average questionnaires´ results 4th A CEIP 14001682

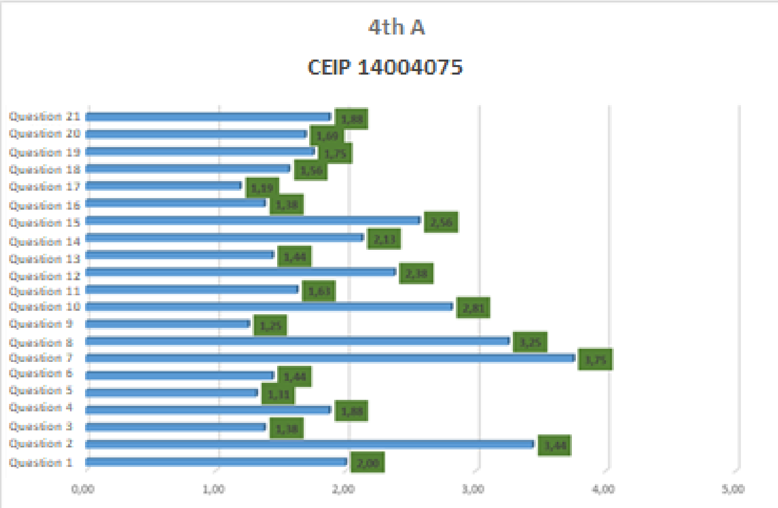

Figure 4. Average questionnaires´ results 4th A CEIP 14004075

Discussion and Conclusion

The aim of this study is to investigate several aspects regarding the acquisition of a foreign language, especially the attitude and motivation of primary school students towards science subjects within the bilingual program. It has been analysed through a questionnaire. Referring to the research questions, students feel motivated and also have a positive attitude towards the bilingual program at school. Likewise, they value learning different subjects through English due to their future possibilities. However, sometimes, it is complicated for them to understand the teacher using the L2. As it can be observed in Q1, some students have problems when using the L2. It is advisable to make students feel comfortable listening and speaking to the foreign language; they need to be confident when doing so. However, referring to Q4, although they got some problems understanding the L2, the majority of students show a great interest in learning English. This result reveals that attitude is decisive and is related to the interest of the students in acquiring a second language. Focusing on the role of the parents, as Q20 reveals, they are a great support, which makes students feel confident learning English. According to Kara (2009), attitude has a key influence on language learning acquisition as well as on the students´ behaviour.

The results show that attitude and motivation towards the bilingual program at school are high. Having an assertive attitude implies great efficiency when learning English, due to the fact that students´ behaviour and attitude are related to language learning (Kara, 2009). However, question number 8 "I feel shy when I have to take part in the lesson" is a negative result which needs to be tackled because English is a language for communication.

Respectively, the number of subjects offered in CEIP 14001682 (science, music and physical education) may have had an influence on these answers, as more bilingual subjects are offered here compared to the CEIP 14004075 (science). Moreover, attention to diversity should also be considered: not all students have the same level of English. In relation to the comparison between the institutions and the subjects which are offered by them, the most relevant aspect can be found in question number 8, where students hesitate when the teacher uses the target language. Additionally, if a comparison is made among the four classes, it can be said that the results are similar among all classes. In question number 15 all learners hesitate. These questions refer to the use of the second language on the part of the teacher and teaching more subjects within the bilingual program at school. Nevertheless, students from the CEIP 14004075 seem to be more in favour of having more subjects within the bilingual program. In light of the results, once again, dealing with attention to diversity is a must in order to solve this dilemma.

With regards to the research question about the motivation of students towards the bilingual program at school, it can be said that they feel motivated: not only due to the fact that they enjoy lessons and eagerly wait for them, but also because they want to learn English as they think it is useful for their future. Secondly, when focusing on the bilingual program in relation to the improvement of their attitude towards English, it should be noted that the bilingual program is helping with this issue: it is improving their attitude towards English because they are constantly working with English and this makes the children happy. Nevertheless, the existence of too great a number of subjects in English can affect their attitude negatively. Thirdly, when students are asked about the importance of learning English, the answer is positive as they view learning English as significant. They value English as a useful tool for the future. However, students do not feel entirely confident when they are not able to understand some concepts in the second language. They sometimes need the teacher to repeat the concepts in Spanish. The results show that students "agree completely" or "agree" with the questions.

Similarly, Dörnyey (1998) argued that motivation affects the success of learners. In order to maintain positive attitudes and motivation among students, the support of the mentor is important. Another suitable pedagogical support is scaffolding. According to David Wood (1988) scaffolding is behaviour intended to tutor, which is contingent, collaborative and interactive. In addition, it is important to consider the different pace of learning and diversity of students (p.96). In light of these considerations, we should be aware of the fact that students' progress might be improved by promoting cooperative or group work and by using ICT (Information and Communications Technology). Some researchers (Gere, 1987; Lawrence & Sommers, 1996; Nystrand, Gamoran, & Heck, 1993) support the assumption that relying on peers and only receiving feedback from teachers creates autonomy in students and independence from teachers, as well as facilitating the deep cognitive processing of the contents they are dealing with.

On the basis of the results, students demonstrate a positive attitude and motivation towards science subjects within the bilingual program. The findings also show that paying attention to diversity is an essential factor that requires more attention. Further research is needed in order to take account of diversity in non-linguistic programmes in order to improve the understanding of language attainment in bilingual programs.

References

Broughton, G., Brumfit, C., Pincas, A., & Wilde, R. D. (2002). *Teaching English as a Foreign Language*. Routledge Education Books.

Campbell, D. J., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N.L. Gage (Ed.), *Handbook of Research on Teaching*. Chicago: Rand McNally.

Coyle, D. (2010). Foreword. In D. Lasagabaster and Y. Ruiz de Zarobe (Eds.), *CLIL in Spain: Implementation, results and teacher training* (pp. vii-viii). Newcastle upon Tyne: Cambridge Scholars.

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum Publishing Co.

Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Language Teaching, 31*(03), 117-135. doi: 10.1017/s026144480001315x

García, O. (2009). *Bilingual Education in the 21st Century: A Global Perspective.* West Sussex, UK: Wiley-Blackwell.

Gardner, R. (1985). *Social psychology and second language learning. The role of attitudes and motivation*. London: Edward Arnold.

Gardner, R., & Lambert, W. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House.

Gere, A. (1987). *Writing groups: History, theory, and implications*. Carbondale: Southern Illinois University Press.

Kara, A. (2009). The Effect of a 'Learning Theories' Unit on Students' Attitudes towards Learning. *Australian Journal of Teacher Education, 34(3)*, 100-113. doi: 10.14221/ajte.2009v34n3.5

Latchanna, G., & Dagnew, A. (2009). Attitude of teachers towards the use of active learning methods. *E-journal of All India Association for Educational Research, 21*(1).

Lawrence, S., & Sommers, E. (1996). From the park bench to the (writing) workshop table: Encouraging collaboration among inexperienced writers. *Teaching English in the Two-Year College*, *23*, 101-111.

Lennartsson, F. (2008). *Students' motivation and attitudes towards learning a second language: British and Swedish students' points of view*. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:vxu:diva-2571.

Maljers, A., Marsh, D., & Wolff, D. (Eds.). (2007). *Windows on CLIL: Content and Language Integrated Learning in the European Spotlight*. The Hague: European Platform for Dutch Education.

Mehisto, P. (2012). *Excellence in bilingual education: A guide for school principals*. Cambridge: Cambridge University Press.

Montano, D. E., & Kasprzyk, D. (2008). Theory of reasoned action, theory of planned behaviour, and the integrated behavioural model. In K. Glanz, B. Rimer & K. Viswanath (Eds.), *Health behavior and health education: Theory, research, and practice* (pp. 67-96). San Francisco, CA: Jossey-Bass.

Noels, K. A. (2001). Learning Spanish as a second language: Learners' orientations and perceptions of their teachers' communication style. *Language Learning, 51*, 107–144.

Noels, K. A., Pelletier, L. G., Clément, R., & Vallerand, R. J. (2003). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning*, *50*, 57–85. https://doi.org/10.1111/1467-9922.53223

Nystrand, M., Gamoran, A., & Heck, M. (1993). Using small groups for response to and thinking about literature. *English Journal*, *82*, 14-22. doi: 10.2307/820670

Pavón, V., & Ávila, J. (Eds.). (2009). *Aplicaciones didácticas para la enseñanza integrada de lengua y contenidos (AICLE/CLIL/ÉMILE*). Sevilla: Consejería de Educación de la Junta de Andalucía-Universidad de Córdoba.

Pérez-Cañado, M. L. (2012). CLIL research in Europe: Past, present, and future. *International Journal of Bilingual Education and Bilingualism*, *15*(3), 315–341. doi:10.1080/13670050.2011.630064

Pérez-Paredes, P., y Martínez-Sánchez, F. (2000-2001). A Spanish version of the foreign language classroom anxiety scale: Revisiting Aida's factor analysis. *Revista Española de Lingüística Aplicada RESLA, 14*, 337-352.

Reid, N. (2003). *Getting started in pedagogical research in the physical sciences*. Hull: LTSN Physical Sciences Centre, University of Hull.

Victori, M., & Lockhart, W. (1995). Enhancing metacognition in self-directed language learning. *System, 23,* 223-234. http://dx.doi.org/10.1016/0346-251X(95)00010-H

Wood, D. J. (1988). *How Children Think and Learn*. Oxford: Blackwell.

Beatriz Calderón Jurado is graduated in the Master's Degree in Advanced English Studies (Cognitive Linguistics/Literature) and Bilingual Education. University of Córdoba, Spain.

Cristina Morilla García is Lecturer in the Department of English and German Philology, University of Cordoba, Spain.

Contact Address: Department of English and German Philology, University of Cordoba, Plaza Cardenal Salazar, 3 Facultad de Filosofía y Letras, Córdoba, Spain. Email: cristmogar@hotmail.com

Aprendizaje e Interacciones en el Aula [Learning and Classroom Interactions]

Elisabeth Torras Gómez[1]

1) University of Barcelona, Spain

Date of publication: October 24th, 2018
Edition period: October 2018 - February 2019

To cite this article: Torras Gómez, E. (2018). [Book Review of *Aprendizaje e interacciones en el aula* by N. Mercer, L. Hargreaves, & R. García-Carrión]. *International Journal of Educational Psychology*, *7*(3), 343-345. doi: 10.17583/ijep.2018.3799

To link this article: http://dx.doi.org/10.17583/ijep.2018.3799

2018 Hipatia Press
ISSN: 2014-3591
DOI: 10.17583/ijep.2018.3799

IJEP – International Journal of Educational Psychology, Vol. 7 No. 3 October 2018 pp. 343-345

Review

N. Mercer, L. Hargreaves, & R. García-Carrión (2016). *Aprendizaje e interacciones en el aula*. Barcelona: Hipatia editorial.

In 2015, the first International Scientific Congress on Learning and Interactions in the Classroom (LeCI) was organized by the Community of Research on Excellence for All, CREA, and the University of Valencia. The aim of this congress was to share the latest advances in scientific research regarding how people learn and develop all human dimensions to their full extent. Three renowned experts in the field took part in the congress: Neil Mercer[1], Linda Hargreaves[2] and Rocío García-Carrión[3]. The results on educational research carried out by the scientific community and that were shared in this congress show how interaction and dialogue are two key elements in the development of the human capacities. It reveals that language is a key factor of learning because it has the power to incorporate the intersubjective dimension to the learning process, as it considers both the knowledge of others and the knowledge built with others. Therefore, it sets the focus on the elements that should be present in any class in order to maximize student interaction and foster the collective creation of meaning. The chapters which follow the introduction are then devoted to the exploration of these aspects.

In the first chapter, Mercer starts from Vygotsky's conception of language as a tool that allows multiple individuals to combine their intelligence and collectively create meaning. Through scientific studies, Mercer explains how in classrooms where teachers establish some ground rules to foster "exploratory speech", which is based on exchanging founded viewpoints and critically analyzing them, students learn to collectively think and create meaning, developing higher levels of reasoning skills. Some students do not experience these dialogues at home, so the class may be the second and last chance to develop the valuable skills they promote, hence the importance of fostering exploratory speech in learning.

In the second chapter, Hargreaves explains how, traditionally, teachers have monopolized the speech in teaching and that students barely interact in class. In this context, she refers to "social pedagogy" and, citing several research projects, exposes how when students work in heterogeneous small groups to solve challenging tasks and follow the ground rules that guide group work, the quality and quantity of interactions is maximized, and students reach higher levels of learning. The teacher's role is then to foster and moderate those interactions and give support.

The third chapter covers García-Carrion's conference regarding scientific findings on Dialogic Literary Gatherings (DLG) as a tool to foster children participation in the classroom. In DLG students interact over the literary classics they are reading in an egalitarian dialogue, in which all contribution must be supported by arguments based on validity claims that are critically reviewed by everyone. This allows students to collectively create meaning around the valuable issues covered in the Classics and maximize the level of what they internalize. Garcia-Carrion explains how DLG have transformed the traditional patterns of classroom participation and changed students' and teachers' perceptions of what they can learn, scientifically proving to raise academic success, to improve the communicative skills and to ameliorate the relationships between classmates, thanks to the promotion of solidarity and the understanding of others.

In the last chapter the symposium which took place at the end of the congress is presented. The dialogic turn in education is acknowledged and relevant issues are addressed by the researchers, such as the necessary elements for quality group work, the evaluation of the quality of interactions, the importance of heterogeneous student aggrupation, the misinterpretations around the "Multiples Intelligences" theory and its educational risks or the need to foster "exploratory speech" when group work is developed using ICT.

Through the presentation of multiple scientific researches, this book shows the benefits of an educational approach that acknowledges the potential of classroom interactions for students' learning and development, not only regarding their academic skills, but also their emotional and social ones. The contributions of this publication are especially relevant when fighting to eliminate occurrence-based classroom practices, as they show the effectiveness of evidence-based interventions in the promotion of success. Thanks to this book, more policy-makers and teachers will be able to reorient their practices so more children get their "second chance" by benefiting from the latest advances in educational research.

Elisabeth Torras Gómez
University of Barcelona
etorras@ub.edu

Notes

[1] Emeritus Professor of Education at Cambridge University, Life Fellow of Hughes Hall and Director at Oracy@Cambridge

[2] Emeritus Professor of Classroom learning and Pedagogy at Cambridge University

[3] Ramon y Cajal Fellow at Universitat de Lleida. Marie Curie Fellow at the University of Cambridge (2013-2015).