Articles

Reviews

Including Psychology in Inclusive Pedagogy: Enriching the Dialogue?

Ruth Kershner[1]

1) University of Cambridge

PLEASE SCROLL DOWN FOR ARTICLE

Including Psychology in Inclusive Pedagogy: Enriching the Dialogue?

Ruth Kershner
University of Cambridge

Abstract

Inclusive education is a complex field of study and practice that requires good communication and dialogue between all involved. Psychology has to some extent been marginalised in these educational dialogues. This is, in part, due to psychology's perceived heritage in the standardised testing that has been used to support the educational segregation of certain individuals and groups of students. Some have also expressed fundamental doubts about the prospects of investigating human experience and education through 'scientific' method in psychology. In this paper I discuss the relationship between inclusive education, dialogue and psychology, with a focus on the dialogic aspects of inclusive classroom pedagogy. I analyse how a group of eight early career primary (elementary) school teachers in England talk about inclusive pedagogy at the start their involvement in a one-year research project on this topic. Their conversation suggests the strong presence of psychological thinking, alongside the teachers' other references to classroom practice, children's rights and social identities. Conclusions are drawn about the need to include the heterogeneous field of psychology in the continuing dialogues of inclusive education, while also considering new forms of psychology for inclusive education.

Keywords: Psychology; inclusive education; pedagogy; dialogue; teachers' talk; primary (elementary) education

Hipatia Press
www.hipatiapress.com

Incluyendo la Psicología en la Pedagogía Inclusiva: ¿Un Diálogo Enriquecido?

Ruth Kershner
University of Cambridge

Resumen

La inclusión educativa es un campo de estudio complejo, que requiere de una buena comunicación y diálogo de todos los implicados. Hasta cierto punto, la psicología ha sido marginalizada de estos diálogos educativos. Esto se debe, en parte, a la percepción de que la tradición de evaluación psicológica estandarizada ha servido a la segregación educacional de ciertos individuos y grupos. Hay quienes dudan de la agenda de investigación que pretende analizar la experiencia y educación humanas mediante el método "científico" en psicología. En este artículo, discuto la relación entre inclusión educativa, diálogo y psicología, poniendo el foco en los aspectos dialógicos de una pedagogía inclusiva. En el estudio participó un grupo de ocho profesores en sus primeros años de ejercicio docente en escuelas primarias de Inglaterra. Específicamente, analizo la forma en que se los docentes se refieren a la pedagogía inclusiva al comienzo de un proyecto de investigación en el área, de un año de duración. Sus conversaciones sugieren una fuerte presencia de pensamiento psicológico, además de referencias a prácticas de aula, derechos del niño e identidades sociales. Las conclusiones apuntan a la necesidad de incluir el heterogéneo campo de la psicología en los diálogos acerca de inclusión educativa. A su vez, aparece la necesidad de considerar nuevas formas de psicología para la inclusión educativa.

Palabras clave: psicología; educación inclusiva; pedagogía; diálogo; habla de los docentes; educación primaria.

Hipatia Press
www.hipatiapress.com

I nclusive education is so diverse and complex that those who are engaged in research, practice and policy development may only glimpse others at a distance, moving in all directions. Some may be trekking purposefully towards the same destination as ourselves, while others appear to be making a totally different journey. In developing inclusive education we need to find a way to communicate with anyone who is concerned with the rights and interests of all students learning in very different social and educational contexts. This can sometimes be difficult when people's aims, values, knowledge and beliefs about inclusion can vary so widely.

The central argument of this paper is that inclusive education is achieved in dialogue with others. This is not just a case of finding ways to communicate effectively, although that is important. It is also a matter of engaging actively in the often challenging process of hearing different voices and seeing different perspectives without necessarily reaching synthesis or agreement (Wegerif, 2008). Inclusive education holds intrinsic tensions and dilemmas that may not reach final resolution, although some sort of coherent view is required for ethical and practical reasons (Norwich, 2014). Compromise may be essential when individual human costs and benefits are weighed up, but active dialogues need to continue in the system itself.

Psychological thinking and practice is commonly marginalised, ignored or rejected in dialogues about educational inclusion. There are reasons for this, not least because psychological research has been negatively implicated in the practices of individual testing that have been used to justify certain students' educational separation and exclusion (Croizet, 2013; Greenstein, 2016; Thomas & Loxley, 2001). Thomas (2014) is concerned that scientific knowledge may be valued over personal knowledge when questions arise about why some children are not succeeding at school. Bridges (2013) argues further that the scientific basis of much psychology can only ever lead to a partial understanding of human experience and education.

Yet it seems odd to squeeze psychology out of research on inclusive education. Inclusive education is inevitably concerned with a whole of range of topics that have been investigated under the umbrella of 'psychology', including:

- the experiences of students and teachers in school
- student identity, motivation and learning
- classroom communication and relationships
- school, home and community links
- approaches to educational assessment
- teachers' pedagogical knowledge, practice and professional learning
- multi-professional communication and teamwork
- the organisational operation of schools
- and so on ….

In this paper I am concerned with how a psychological perspective, broadly defined, may contribute positively to the dialogic engagements of inclusive education. This includes some implications for valuing new forms of psychological thought and methodology.

Researching Inclusive Education

Recent studies of inclusive education have reiterated the need to take stock of the field and examine what is actually happening in different practice contexts. Smyth et al (2014), for instance, adopt an international comparative perspective. They trace the 'implementation trajectories' of four European countries (Ireland, Austria, Spain and Czech Republic) moving towards more inclusive education systems within a common international UN and European policy environment. They conclude that

> …(w)hile there is apparently broad agreement at an international level about what inclusive learning environments should look like, there is considerably less agreement about how this can be achieved at national and local community level. The range of legacy interests, pressures and priorities operational in individual education systems is inevitably shaping the manifestation of enabling legislation as well as of provision within schools. (p.442).

This apparent lack of consistency in developing inclusive learning environments is not entirely surprising, given the competing pressures

applying in local and national educational systems. There is a question of system capacity here. Ainscow et al (2016) analyse how English primary schools and teachers can respond to increasingly diverse populations of learners. They comment on the relevance of three interlinked sets of factors that apply *within* schools, *between* schools and *beyond* schools, bearing in mind local demographics, economics, cultures and histories. This points to the need for those in school to share practices with each other, developing a common language to do so, and for schools to collaborate more widely with each other, with community partners and researchers. Ideally it tips the balance away from generic 'what works' approaches towards the knowledge that is grown in local school contexts.

This type of recent of work on inclusion not only raises questions about the inherent power relationships and other constraints in complex educational systems, it also highlights the conceptual challenges that can hinder communication and dialogue. Researchers have long acknowledged the conceptual difficulties in defining inclusion. In a recent review, Göransson & Nilholm (2014) identify four different types of definition: the physical placement of children with identified disabilities or in need of special support in general education classrooms; meeting the social/academic needs of these identified pupils; meeting the needs of *all* pupils; and the creation of school and classroom communities which are participatory, equitable and valuing of diversity. In conclusion, Göransson & Nilholm remark from philosophical and political perspectives that: '…the definitional problems indicate differences in beliefs about what schools can and should accomplish. This brings the question of power into the analysis. Who should decide what version of inclusion should be the goal of schooling?' (p.275)

In commenting on this work from his own experience of conducting an earlier systematic review on how schools facilitate all students' participation, Dyson (2014) turns to the intrinsic difficulties of conducting research using established methodologies when it remains unclear that inclusion can actually be studied in this way. His concern is that: '…for many researchers, inclusive education is not a set of practices whose effects can be evaluated, but is a principle (or, more accurately, a set of principles) which is embodied in different ways in different contexts.' (p.282)

Dyson sees the consequences in research that is 'descriptive, celebratory and exhortatory' rather than convincingly evaluative. The associated danger, in his view, is that inclusion research is inevitably limited in scope and potentially stranded on the moral high ground without influence on practitioners and policy-makers. This is where authentic dialogue is required. The creative acknowledgement of a dialogic gap or difference in perspectives is a prime source of meaning and creativity in a complex and often problematic situations. As Wegerif (2007:28) puts it: …(it) indicates a challenging direction of development for individuals and society towards a greater capacity for creative thinking and a greater capacity for learning to learn, intimately linked to an ethics of openness to the other'.

Psychologically this requires the co-ordination of different perspectives, not just between people in conversation but in our own thinking. Fernyhough (2016) remarks on this in his discussion of 'inner speech' and dialogic learning from a cognitive perspective: 'Thinking is a dialogue, and human cognition retains many of the powers of a conversation between different points of view' (p.98).

Others agree that conversation is important as a basis for creativity and problem-solving:

> …when people of any age are working together to create new ideas and understandings, (t)hey use talk and joint activity to create a shared resource of ideas that can be jointly considered, and a framework for collected working that will enable their work to progress. (Littleton & Mercer, 2013, p. 110)

I have deliberately presented this brief section with several quotes from relevant literature, representing a (selective) range of perspectives. These and other written voices influence our own thinking, and we in turn bring new ideas to our conversations and writing about inclusive education. There is an inevitable selection process involved in the attention and value given to different speakers and forms of knowledge - we would otherwise be overwhelmed with a cacophony of voices. We can, however, ask if the selection of different points of view is random or systematic within our own thinking and in the more formal processes of research, policy and practice – i.e. what forms of knowledge are valued and used, where and by whom?

Dialogues in and about Inclusive Education

It is hard to imagine an inclusive educational system without people's willingness to consider other perspectives and engage in productive dialogue, including the political negotiations involved in setting political priorities and education budgets. Inclusion depends on people's capacity and willingness to communicate across boundaries of geography, language, professional priorities and personal concerns.

A wealth of research has turned towards a broadly dialogic perspective on inclusion, with interests ranging from the uses of talk and dialogue as means of involving marginalised communities in education to the means of classroom learning and teaching for all students. This body of work represents rather different perspectives on dialogue, informed by theoretical debates such as those between followers of Vygotsky and Bakhtin (Wegerif, 2008). Some studies focus on the social and educational value of talk and other forms of communication, generally following a social constructivist model of joint activity and learning. Others look to the creation of meaning within the contrasting discourses and alternative perspectives of human relations. To give some examples: Flecha (2011) discusses a 'dialogic sociology of education' that emphasises the role of communities and human agency in challenging unequal structures and practices in society. Everyday processes are discussed and developed jointly between researchers, practitioners, students, parents and other community members to implement 'successful educational actions' that help to overcome social inequalities. One such action is seen in the adoption of heterogenous classroom groupings designed to promote social interaction, dialogue and learning for all students, including vulnerable minority ethnic populations, with the support of adult community members (Valls & Kyriakides, 2013). Classroom conversation has been acknowledged as key to inclusion in settings where some children may be identified with special educational needs and disabilities (Berry & Englert, 2005; Ní Bhroin, 2013), while Rajala, Hilppö & Lipponen (2012) examine whether a form of 'exploratory' talk known to support higher order thinking is itself equitable and inclusive of different students. Studies have also been conducted using interviews and discussions to gain better understanding of students' experiences of inclusion both during and after

their school experience (Adderley *et al,* 2014; Diez, 2010; Tetler & Baltzer, 2011).

Creating space for professional dialogue has been seen as central to the development of inclusive schools (Howes, Grimes & Shohel, 2011) and pedagogical innovation (Moate, 2014). Collaborative and inclusive practices in schools have been promoted through dialogue, including a goal-directed use of Socratic method designed to challenge teachers' thinking and bring in new practices (Tragoulia & Strogilos, 2013). It has been suggested that a shared pedagogical vision is important for students and others to feel that they belong to an inclusive school community (Hazel & Allen, 2013), but one of the features of a (Bakhtinian) dialogic understanding of school development is that it does not pre-suppose consensus in school about provision required for students experiencing difficulty (Skidmore, 1999).

We can see from this brief selection of research that communication may not only be required to argue <u>for</u> the development of inclusive education, but communication and different forms of dialogue are ideally embedded <u>in</u> the experience of inclusive education. This is particularly evident when describing what inclusive education may look like in classroom practice with reference to student learning. Sheehy (2013) refers to a review of pedagogies and outcomes for the academic and social inclusion of pupils identified with special educational needs in mainstream classrooms (Sheehy & Rix, et al. 2009), identifying the pedagogic features of an inclusive classroom as:

1. Social engagement being intrinsic to the pedagogy
2. Flexible modes of representing activities
3. Progressing scaffolding of classroom activities
4. Authenticity of classroom activities
5. Pedagogic community

Sheehy (2013, p. 242) summarises:

> …this refers to the teacher facilitating cooperative group work, using a variety of representations of problems to present and discuss issues… (Pupils) engage gradually with concepts and develop the skills they need….(T)he teacher is part of pedagogic community. They are supported by, and contribute to, a group who

> have a shared view of what they are teaching and why they are teaching it in a particular way.

This belief in the importance of classroom communication is supported in other discussions of inclusive pedagogy, especially those that adopt a social constructivist perspective on the collective experience of inclusive classroom learning. Situated and distributed understandings of cognition, together with opportunities for multimodal communication, point towards the incorporation of purposeful, real-world classroom activities to engage students and teachers in valuing each other's experience, constructing knowledge together and developing higher level cognitive skills (Kershner, 2009). Reviews of educational programmes designed for cultural and/or linguistic minority students come to similar conclusions, as we see in Tharp & Dalton's (2008) account of the universal features of classrooms that promote educational success for diverse and at-risk student populations. Their standards to meet in classroom teaching include the following, with the prospect of more to come:

1. Teachers and students producing together
2. Developing language and literacy across the curriculum
3. Making meaning – connecting school to students' lives
4. Teaching complex thinking
5. Teaching through instructional conversation

Tharp & Dalton (2008) remark that while the effects of this pedagogy may be directly attributed to the means of assisting students' performance and promoting development, there are different strands of theoretical thinking that offer further support. For instance, they suggest that cultural-historical-activity theory's focus on relating the personal to the cultural with a development orientation can be placed alongside a cognitive science perspective on the efficacy of instructional dialogue and contextualisation for prompting cognitive processing, improving conceptual retention and reducing cognitive load.

In relation to these views of inclusive classroom practice we can further

acknowledge the essential social and emotional conditions of learning (Noddings, 1992) as well as the importance of teachers' beliefs and strategies for teaching all students by responding respectfully to student difference and rejecting deterministic beliefs about ability to learn (Florian and Black-Hawkins, 2011). Thus, the essential notion of 'pedagogy' within inclusive pedagogy extends beyond the overt techniques and knowledge of teaching to encompass the classroom relationships and educational purposes that support all children's learning. A teacher's concern for the well-being and flourishing of individual learners in class is accompanied by similar concern for the whole class group. Inevitably these perspectives do not always align or balance within and beyond each school but, as mentioned above, consensus may not be necessary for a school to move forward.

While the communication between students and teachers in inclusive classrooms and schools is well-established, less attention has been given directly to how people talk about the concept of inclusive education itself. We do not know a great deal about what may support or hinder productive dialogue about inclusion when attitudes and beliefs can vary so widely. This is not due to lack of attention to people's thoughts on the topic, particularly with regard to teachers. There is great research interest in teachers' *attitudes* to inclusion, for instance. A search of the British Education Index combining 'attitude' and 'inclusion' and 'teacher' resulted in 276 recent references at the time of writing this paper, with empirical studies reported from across the world. In comparison there seems to be less work on what happens when teachers and others discuss their various ideas and concerns about inclusion.

In order to explore the ways in which dialogues about inclusion may actually operate between people, I will turn next to an example of how teachers who are interested in inclusion talked about it when focusing on their pedagogical practice. What comes to the minds of teachers in such conversations?

Talking About Inclusive Pedagogy

This discussion below occurred during a recent project that I have been involved in with my Faculty co-investigator Dr. Kristine Black-Hawkins. A group of eight primary (elementary) teachers in early career (within 1-3 years of qualifying) had joined a research project focusing on the

development of inclusive pedagogy. The year-long project included four Faculty-based workshops for the teachers, as well as teachers' classroom-based inquiries and wider school-based discussions with the teachers, senior leaders and pupils (Black-Hawkins and Kershner, in preparation). At the first workshop the following discussion activity, which was devised and led by Kristine Black-Hawkins was set up both as a social ice-breaker and as a means of activating the teachers' thinking and talk about inclusive pedagogy. The teachers were asked to work in pairs to respond to four key words written on large sheets of paper. These words were selected to represent possible aspects or components of inclusive classroom learning that could be meaningful to the teachers: BELONGING, LEARNING, DIVERSITY, PARTICIPATION. (The term 'inclusion' was deliberately avoided at this point, partly in order to prompt and trace alternative ways of thinking from the start of the project.) The tables were set out so that each pair started with one of the words and then moved the next in the same order on the facilitator's signal. This continued until all the groups had made written comments in response to all four key words and then returned to their starting place for a final discussion. Each round of conversation was just 2-3 minutes long and the whole activity was completed within 20 minutes. The conversations were audio recorded and then transcribed using an 'intelligent verbatim' style that omits repetitions and filler words.

There are different ways in which this sort of conversation can be analysed, according to different purposes (Littleton and Mercer, 2013). A *linguistic* analysis may focus on conversational acts such as questioning, while a *psychological* analysis could be concerned with the communicative relationship and thinking (e.g. collective reasoning, response to other speakers, and discussion of particular topics). A *cultural* level could include the conversational 'ground rules' and the communicative principles that are valued in that context (e.g. clarity, respect for others, and use of evidence). My focus here is primarily at the psychological level, giving attention to the teachers' collective thinking, as represented in their conversations. The research questions are open: How did these early career teachers respond to the key words in their conversations? Are there any indications of 'psychological' thinking?

Here is an account of the five short conversations around the key word of 'BELONGING'. All are focused on ideas cumulatively written up on the sheet:

Round 1: The first contribution comes from Teacher 1: 'So, feeling like you're in a safe place'. The discussion then continues between the two teachers in an affirmative and personal way:

> Teacher 2: Yes, feeling like you're being in a safe place. That everyone matters, everyone belongs, not just some people.
> Teacher 1: Feeling like it's a family, almost
> Teacher 2: Yes, I definitely agree with that. Actually, my headmistress came in and said, 'Oh, it's like a little family in here.' She said that to my class.
> Teacher 1: Sharing happiness, sharing things.
> Teacher 2: Yes, that's good, sharing. If you feel excluded that someone is actually going to be there, that you have support.

This part of the conversation focuses almost entirely on feelings, sharing, supporting and ensuring that people are involved. It is only towards the end of this first round that mention is made of academic learning, and the focus here is on ensuring that learning tasks are appropriate for the children. The teachers acknowledge that this is hard, though, especially if this leads to dividing children into groups based on their attainment:

> Teacher 1: … I'm just thinking that there are some children who feel that they are in the wrong. So, we set them for maths and they feel like, because they're in the bottom group, they don't feel that they belong there. So, academically they might belong there, but they feel that they're in the wrong place

Round 2: The next pair of teachers agrees that 'the safe place is fundamental' and 'paramount'. As with the first pair the conversation turns to their own classroom experience and the strategies they use to enhance children's feelings of safety and belonging.

> Teacher 3: I would definitely say in my classroom … the 'safe place' is fundamental.

> Teacher 4: Paramount, yes.
> Teacher 3: Because in my room I know from last year that was not the case for my children.
> Teacher 4: Really?
> Teacher 3: This year straightaway the anxieties that are coming in, don't feel safe in their room. So, really trying to support that feeling of belonging and scaffold them into feeling they belong in that room.

Their conversation turns to the children's 'ownership' of the classroom and how this can support their increasing independence as learners.

> Teacher 4: Yes, I think that sounds good to me. I know it sounds like your jargonistic way of approaching it, but the ownership of the room, so they get to change the actual environment themselves. So, when we develop the environment, we put in new things like role play areas, they have a big say in that. As a result, it means that it's their room and it's an environment that they're not surprised by coming in, only on a special day where we might have done something for them. But it's actually gradually becoming theirs and theirs.
> Teacher 3: Building onto that, if they feel they belong and that classroom belongs to them, they become more independent as learners.

The two teachers then read through the previous pair's notes, agreeing with all that is written. They decide to add 'child-initiated' and 'being welcome', which '...is something that we do quite well'.

The two teachers then begin to articulate their principles, although these do not yet emerge clearly in their conversation

> Teacher 4: It's a concept, a kind of philosophy for me, anyway, beyond the things. I can see how the work comes into it, but, for me, it's very much your approach.
> Teacher 3: Yes, psychological thing as well, because it can become, if they feel you're in control of the whole space, it then goes on a hierarchy as well, where they're totally on balance.

This leads them to discuss question of 'voice' and 'control'. They touch on the question of who retains the control in class, agreeing that this goes to the teacher in the end. They agree that the concepts on the sheet are 'nicely linked'.

One also remarks further on the 'sharing' notion already present on the sheet: Teacher 3: 'Yes, and I think with sharing it could be sharing space as well and sharing knowledge'.

Round 3: The arrival of the third pair adds new reference to 'community' and '...feeling that your ideas are valued as much as anyone else's'.

> Teacher 5: Belonging? A sense of belonging? So, feeling as though you're part of the class.
> Teacher 6: Part of the community.
> Teacher 5: And part of the wider community. That's true. Feeling that your ideas are valued as much as anybody else's.

They then turn their conversation to religion:

> Teacher 5: Also, in terms of, if you're looking from an RE (*religious education*) perspective, your sense of belonging and how it differs, perhaps, from other people, how they see belonging, like belonging to a religion. I don't know. Not just your community, belonging to a faith.

Their brief discussion of faith groups leads them to consider any groups and the implications for children '...who don't have friends, who feel like outsiders, because of whatever reason'. They continue thinking about belonging in terms of whether the feeling of belonging arises '...when you share similar interests....or when you're taken seriously...that you matter..and that you're cared for', bringing in an example from their own classroom:

> Teacher 5: Yes. I think children want to belong, in the sense that they want to have similar things that other children are having and like similar things, maybe, links to friends. For instance, one girl in our class decided she wanted to wear a different coloured pair of tights, because her friend had them. She wanted to feel like she belonged in that room.

Teacher 6: Yes. They do need to feel secure, then, don't they?
Teacher 5: Yes, maybe.

They conclude their conversation reflecting on the questions of security and whether children want to stand out', noting in the end that

Teacher 6: It's not often a child really wants to stand out, is it?
Teacher 5: Similar interests. Yes, that's true.
Teacher 6: It's hard.
Teacher 5: But, then, is that something that we promote as teachers, that everybody does one answer, maybe sometimes? I don't know. Maybe that's promoted that everybody needs to be doing a similar thing, like the behaviour management. So, maybe that's belonging.

Round 4: The fourth pair of teachers are immediately attracted to what is already written down about 'feeling safe', and one of them adds the word 'nurturing'. They then strike out in two directions: one teacher follows a line of thinking about 'feeling that you can speak your mind' while the other heads in the direction of learning (perhaps reflecting their conversation about another key word encountered in previous rounds)

Teacher 7: I think if you belong somewhere you feel like you can speak your mind.
Teacher 8: Yes, exactly. And you have to feel that you can belong before you can participate or learn.

These teachers then continue to read and comment on the nearly full sheet of ideas, adding their ideas about the need for respect and sensitivity to children's different backgrounds. They take this further in reflecting on the need to be sensitive when children are '... taken out of class to do things'. Teacher 7 extends this point to reflect on the difficulties of integration between environments. The teachers conclude by building on previous ideas with a sense of completion on the sheet, adding references to peers, family and friends. They check that mention has been of 'thinking you're valued' and 'having a voice'. Their final contribution is to extend the general notion of 'family' to consider the involvement of children's parents and the wider

community in school, noting the difficulties and dilemmas of control that can arise:

> Teacher 8: Yes, because in our school we struggle a lot with getting parents in and doing stuff with the community, because there is this big barrier. I guess it's because the parents feel that they don't belong in school, because they've had a negative experience of that. So, they don't really want to come in ever for anything. If they do, they will want to be in and out as soon as they can.
>
> Teacher 7: Sometimes it's difficult because the parents aren't allowed to come in, in the morning, but then we want to be like an open school. So it's that conflict.

Round 5: In the final round the first pair return to the sheet that they started off, and they are asked to select one or two points to feed back to the whole group. They comment first on the ideas that they like, such as 'child-initiated' and 'multiple voices'. They also identify ideas they don't understand such as 'control' and 'hierarchy', which prompts further conversation and co-construction as they try to work out what it could mean:

> Teacher 1: I don't understand this one.
>
> Teacher 2: I know. I don't understand 'control' or 'hierarchy' either.
>
> Teacher 1: I suppose you get the safe place by having control of the classroom as a teacher, because, if the classroom is not under control, then it's not a safe place.
>
> Teacher 2: I guess I don't think teachers can control the classroom. I think teachers can…
>
> Teacher 1: … manage it.
>
> Teacher 2: Well, the children are only, at the end of the day, have control over, really get control children. They choose to control themselves.
>
> Teacher 1: Yes, but that's still through how you set up and manage that.
>
> Teacher 2: Yes, exactly, how I manage it……

In the end they select 'a safe place' as their 'favourite', remembering that

one of them had offered this idea originally. They complete their argument as follows:

> Teacher 1: Yes, exactly, how I manage it. Did you like that, 'a safe place'?
> Teacher 2: I did it.
> Teacher 1: I like that. I think that's probably my favourite one. I'm going to put a star beside it. I like that one. I do. I think that is.
> Teacher 2: Because I think all those other things are possible steps from it. If a child chooses to participate, it's because they feel safe.
> Teacher 1: Yes, absolutely

This series of short conversations about 'belonging' has been presented in detail to demonstrate how rapidly and fluently the teachers built on each other's ideas within and between each short round of conversation. The teachers' talk has largely 'cumulative' features, meaning that they mostly accept other's ideas and elaborate on them in an affirming way. They tend not to adopt the questioning, critical evaluation or challenging approaches typical of the 'exploratory talk' that is associated more strongly with collective reasoning and problem-solving (Littleton and Mercer, 2013). This seems unsurprising given the nature of the activity and its role early in the project. Indeed, Littleton and Mercer remark on the value of cumulative talk in certain stages of joint creative activity (p.58).

The teachers bring different types of knowledge into the conversations, often starting with what seems to be initial free word association and extending to the inclusion of personal anecdotes, with occasional reference to more formal theoretical ideas and terminology. They touch on many psychologically relevant ideas, with particular reference to feeling safe, feelings of belonging, feeling excluded, feeling in the wrong place, children's needs to feel secure, and so on. There are also theoretical propositions, as when Teacher 3 connects children's feelings of belonging with becoming independent learners, Teacher 6 suggests that children do not often really want to stand out, and Teacher 8 says '…you have to feel like you can belong before you can participate and learn'.

In all four key word conversations a great many ideas appeared in a very

short space of time. For instance, approximately 75 different ideas appeared in the conversation about 'LEARNING', without counting repetitions (see Table 1)

Table 1:
Ideas and concepts emerging in teachers' conversations about 'LEARNING'

Rounds of conversation involving teacher pairs (P1-4)	Main ideas and concepts expressed by teachers in conversation about LEARNING (not including repetitions within each round)
Round 1 (P1)	Knowledge; Lessons; Social and academic; Emotional; Progress; Data; Talking; Collaboration; Life skills; Social skills; Friends; Play
Round 2 (P2 – had previously discussed 'belonging'))	Progress; Getting better; Understanding; Lifelong education; Lifelong learning; Common sense; Feeling good about your work; Pride in your work; Enjoyment; Self-efficacy – it's in your control; From each other; Adults and pupils; Non-hierarchical; Motivation
Round 3 (P3 – had previously discussed 'participation' and then 'belonging')	Building together; Safe environment for exploration; Behaviour; Pressure (from management or Ofsted (*i.e. school inspection*); Children knowing boundaries; Understanding their realm; Developing personalities; Self-esteem; Accessing learning; Engaging learning; Enjoyment and challenge
Round 4 (P4 – had previously discussed 'diversity', 'participation' and 'belonging'	In different ways; Participating; Learning in different forms (like learning through play); Social learning; Kinaesthetic learning; Visual; Stages of learning; 'Extend, extend, extend'; Learning support; Measuring learning; Feeling of belonging; Enthusiasm; Learning how to learn; Being independent; Self-regulators; Learning to be a kind person; Us (teachers) learning; Parents learning; Home life learning; Rote vs exploration; Deeper understanding
Round 5 (P1 – i.e. the original pair, returning after discussing 'diversity', 'participation' and 'belonging')	Enjoyment; 'Every child matters'; Life skills; Learning how to learn; Meta-learning; Learning muscles: collaboration, empathy, problem-solving, reasoning and meta-learning; Not spoon-feeding; Independence and autonomy; Knowing how to interact

When considered through a broadly 'psychological' lens some interesting questions may come up about all these conversations. One of the most basic is to ask how the words themselves are being understood and used. It is striking that different topics of conversation emerge in response to each key word. The 'BELONGING' conversation gave precedence throughout to feeling safe, sharing, support and enjoyment, eliciting at least some of the teachers' beliefs about essential relational conditions for learning. Connections to learning are acknowledged, mainly in terms of increasing independence, and the discussion develops towards questions of ownership, voice and control. In contrast, the 'DIVERSITY' key word conversation started with discussion about different religions, cultures, languages, and socio-economic status. Reference is made to classroom learning, but in terms of the relevance of 'ability', 'learning needs' 'gifted and talented'. Final emphasis is given to 'understanding that people are different and unique'. The key word 'PARTICIPATION' prompted initial conversation about active learning in the classroom, choice, decision-making and active participation in democratic school life, and the associated needs for communication and understanding. The lengthiest discussion here was actually about the dilemmas and limits of children's active participation in school. One teacher in an earlier round had described a 'no hands up' policy in her school, indicating that everyone is expected to participate in class discussion chosen randomly by the teacher. A later pair, who had been first in the BELONGING conversation above, comment on this point in Round 4:

> Teacher 2: What do you think about this 'no hands up' thing?
> Teacher 1: What, whether it's participation or whether it's a good idea?
> Teacher 2: Is it participation, because actually isn't the teacher forcing participation? Are the children choosing to participate or is the teacher saying 'I'm making you'? You have a right to silence.

This debate continues for a while, with a final philosophical question: 'Is a classroom a democracy or is it a dictator?'

While I was reflecting on how these teachers responded to each other in conversation, and bearing in mind the psychological focus of this paper, certain words came to my mind, including: 'lexical' and 'priming'. These are not areas of psychological research that I know much about, but I

suspected a vast research field. This led me to search psychological literature (using the database PsycINFO) for research on 'lexical' and 'priming' effects that could possibly offer insight into what happens when people talk about inclusive pedagogy in this way, and I made a quick selection of articles that caught my eye. These very different articles drew my attention to such topics as: the effects of people's beliefs about the person they are talking to (Branigan, et al 2011) and the relevance of prior relationships (Ahnert et al, 2013); the activation of 'real-world' knowledge by specific words (Hare et al, 2009); the different types of relationship between pairs of words (Jones and Golonka, 2012); about effects of relevant knowledge on the originality ideas generated in 'brainstorming' (Rietzchel, Nijstad and Stroebe, 2007); and the evolution of nurses' concepts of hospital hygiene over the course of training (Salès-Wuillemin, et al 2011). This rather random set of studies refers to children and adults in different contexts (several in laboratories), and there are no direct applications to inclusive education. However, the process of searching certainly extended my thinking in ways that I could decide to pursue if they seem helpful to understand how people talk about inclusion, just as other lines of reading could do the same for different purposes.

This section has raised questions about how people can share ideas and come to understand each other in conversation. I have also touched on the knowledge that can be incorporated and developed in dialogue both in direct conversation with others and in virtual dialogue with published work. Both seem relevant to understanding the foundations of constructive and productive dialogue about inclusive education.

Conclusion: Inclusive Education, Dialogue and Psychology

In this paper I have begun to consider whether and how psychological thinking may contribute to the dialogues essential for developing inclusive education, with particular reference to classroom practice and teachers' thinking in inclusive pedagogy. This is intrinsically a dialogic process in that its meaning and practices involve the engagement of different perspectives, and it is likely to need continual rethinking and innovation. When considering the experience of educational change, I would follow Fullan

(2007) in saying that there is no getting round the 'primacy of personal contact' for teachers to develop shared understanding, moral commitment, trust and coping capacities. Teachers '… need to have one-to-one and group opportunities to receive and give help and more simply to *converse* about the meaning of change'. (p.139).

Given the remaining disciplinary 'wars', it may be that the inclusion of psychological thinking in this area is better expressed in terms of including thinking that draws on ideas, information and knowledge that people who define themselves as psychologists have also been concerned with over the years. This could help to acknowledge common interests and avoid disciplinary arguments. Boundaries around 'psychology' can certainly add to the difficulties of applying psychological knowledge in education, not just because of the concerned skepticism about testing and scientific method mentioned at the start of this paper. It can also be a problem if, as (Hick, Kershner and Farrell, 2009, p. 4) suggest with regard to the extensive educational adoption of concepts like 'learning style': 'The educational "usefulness" of psychology comes to be determined by the success of "non-psychologists" in applying snippets of psychological knowledge and procedures that have somehow gained cultural value'.

It is also important to see that psychological ideas can change over time, sometimes with significant shifts in thinking. Sheehy (2013) points out that psychologists themselves are a heterogeneous group, who adopt different discourses and hold different beliefs that have direct influence on building knowledge about inclusive education through research. This can have methodological impact if new types of quantitative and qualitative evidence gain weight and value. There may also be conceptual change, as we see in Bruner's (2012) reflections on the development of his lab-based work on perception. He concludes that we each look at the world in ways that reflect our situations, expectations, cultural orientations and capacities to construct possible worlds that transcend biological constraints:

> '…do our conventional psychological methods of research – the laboratory, the conventional interview, standardized tests, and the rest – do these take such considerations into account? A psychologist can learn a lesson or two from the anthropologist, the sociologist, even the historian. We will never understand human

behavior simply by studying it in vitro or out of context, without taking account of the uneasy historical compromise that exists between the Established and the Possible.' (p.9)

There is of course existing research that establishes the relevance of psychological theory for understanding likely components of educational inclusion, such as Rose & Norwich's (2014) adoption of social psychological theories of group processes and efficacy beliefs in looking at the resolution of dilemmas in inter-professional work, to name just one. It is also useful to refer to different psychological theories of learning, development and individual difference when considering possibilities for assessment, including psychometric, behavioural, developmental, cognitive, constructivist, humanist, ecological and self-focused approaches (Bourke & Mentis, 2014). We can be open to traditional forms and areas of psychological research while also considering new directions for psychology that will be particularly relevant to educational inclusion. This is likely to involve psychologists in adopting critical approaches that start from social justice principles, less defensiveness about the discipline, and more active engagement in interdisciplinary approaches (Dyson and Howes, 2009; Hick, 2009).

In any case there may be no need to put a boundary around different topics or imply disciplinary ownership. The field of inclusive education seems to be a good candidate for interdisciplinary or transdisciplinary approaches to research and practice, and it is encouraging and exciting to imagine that such work in the field of inclusive education too could lead to new forms of research and new ways of thinking (Darbellay, 2014). Klein (2014) remarks on the overlapping discourses of transdisciplinarity, from its emergence in the 1970s, with a concern for 'imaging futures' in the human, social, technical and natural sciences. She draws attention to the transdisciplinary imperatives of transcendence, problem solving and transgression that play out in an eclectic mix of global and individual projects, relating variously to the study of climate change, architecture, poverty and so on. Education in general, and inclusive education specifically, would seem to be thirsty for such initiatives. To take one example, we might look at the conditions for productive dialogue at different and complementary levels of analysis:

- the power structures and social conditions influencing participation in productive dialogue about educational inclusion
- the means of communication, motivations and social relationships that enable and prompt people to engage with each other to develop more just and equitable educational systems
- the use of tools to support dialogue, helping to articulate assumptions, concepts and actions for inclusive pedagogy in particular contexts (e.g. Florian, 2014)

In conclusion, I have argued that educational inclusion requires conversations and dialogic engagement between all involved. I would add that these face to face and written discussions are potentially enriched by the incorporation of knowledge and understanding gained from relevant psychological and transdisciplinary research that is itself inclusive.

References

Adderley, R., Hope, M.A., Hughes, G.C., Jones, L., Messiou, K., & Shaw, P.A. (2014). Exploring inclusive practices in primary schools: focusing on children's voices. *European Journal of Special Needs Education, 30* (1), 106-121. doi:10.1080/08856257.2014.964580

Ahnert, L., Milatz, A., Kappler, G., & Schneiderwind, J. (2013). The impact of teacher-child relationships on child cognitive performance as explored by a priming paradigm. *Developmental Psychology, 49* (3), 554-567. doi:10.1037/a0031283

Ainscow, M., Dyson, A., Hopwood, L., & Thomson, S. (2016). *Primary Schools Responding to Diversity: Barriers and possibilities.* CPRT

Research Surview 8 (new series). York: Cambridge Primary Review Trust.

Berry, R.A.W. & Englert, C.S. (2005). Designing Conversation: Book Discussions in a Primary Inclusion Classroom. *Learning Disability Quarterly, 28 (1)*, 35-58. doi:10.2307/4126972

Bourke, R. & Mentis, M. (2014). Self-assessment as an 'Insider' Lens for Learning and Assessment. In L. Florian (ed) *The SAGE Handbook of Special Education, 2ⁿᵈ Ed.* (pp. 537-552). London: SAGE.

Branigan, H., Pickering, M.J., Pearson, J., McLean, J.F., & Brown, A. (2011). The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers. *Cognition, 121*, 41-57. doi:10.1016/j.cognition.2011.05.011

Bridges, D. (2013). 'The Proper Study of Mankind'? In defence of the Humanities against the exaggerated pretensions of 'scientific' psychology. In P. Smeyers & M. Depaepe (Eds.), *Educational Research: The Attraction of Psychology* (pp. 131-145). Dordrecht: Springer.

Bruner, J. (2012). What Psychology Should Study. *International Journal of Educational Psychology, 1* (1), 5-13. doi:10.4471/ijep.2012.01

Croizet, J-C. (2013) On the fatal attractiveness of psychology: Racism of intelligence in education. In P. Smeyers & M. Depaepe (Eds.), *Educational Research: The Attraction of Psychology* (pp. 33-51). Dordrecht: Springer.

Darbellay, F. (2014). Rethinking inter- and transdisciplinarity: Undisciplined knowledge and the emergence of a new thought style. *Futures, 65,* 163-174. doi:10.1016/j.futures.2014.10.009

Diez, A. M. (2010). School memories of young people with disabilities: an analysis of barriers and aids to inclusion. *Disability and Society, 25* (2), 163-175. doi: 10.1080/09687590903534346

Dyson, A. & Howes, A. (2009). Towards an interdisciplinary research agenda for inclusive education. In P. Hick, R. Kershner & P. Farrell (Eds.), *Psychology for Inclusive Education: New Directions in Theory and Practice* (pp. 153-164). Abingdon: Routledge.

Dyson, A. (2014). A response to Göransson and Nilholm. *European Journal of Special Needs Education, 29* (3), 281-282.
doi:10.1080/08856257.2014.933542

Fernyhough, C. (2016). *The Voices Within: The history and science of how we talk to ourselves*. London: Profile Books/ Wellcome Collection.

Flecha, R. (2011). The dialogic sociology of education. *International Studies in Sociology of Education, 21* (1) 7-20.
doi:10.1080/09620214.2011.543849

Florian, L. & Black-Hawkins, K. (2011) Exploring inclusive pedagogy. *British Educational Research Journal, 37* (5), 813-828.
doi:10.1080/01411926.2010.501096

Florian, L. (2014). What Counts as Evidence of Inclusive Education? *European Journal of Special Needs Education, 29* (3), 286-294.
doi:10.1080/08856257.2014.933551

Fullan, M. (2007). *The New Meaning of Educational Change, 4th ed*. New York, NY: Teachers College Press.

Göransson, K. and Nilholm. C. (2014). Conceptual Diversities and Empirical Shortcomings: A critical analysis of research on inclusive education. *European Journal of Special Needs Education, 29* (3), 265-280. doi:10.1080/08856257.2014.933545

Greenstein, A. (2016). *Radical Inclusive Education: Disability, teaching and the struggles for liberation*. Hove: Routledge.

Hare, M., Jones, M., Thomson, C., Kelly, S., & McRae, K. (2009). Activating event knowledge. *Cognition, 111*, 151-167.
doi:10.1016/j.cognition.2009.01.009

Hazel, C.E. & Allen, W.B. (2013). Creating inclusive communities through pedagogy at three elementary schools *School Effectiveness and School Improvement, 24* (3), 336-356.
doi:10.1080/09243453.2012.692696

Hick, P. (2009) Reframing psychology for learning within social justice agendas. In P. Hick, R. Kershner & P. Farrell (Eds.), *Psychology for Inclusive Education: New Directions in Theory and Practice* (pp. 165-176). Abingdon: Routledge.

Hick, P., Kershner, R. & Farrell, P. (2009). Introduction. In P. Hick, R. Kershner & P. Farrell (Eds.), *Psychology for Inclusive Education:*

New Directions in Theory and Practice (pp. 1-10). Abingdon: Routledge.

Howes, A.J., Grimes, P., & Shohel, M.M.C. (2011). Imagining inclusive teachers: contesting policy assumptions in relation to the development of inclusive practice in schools. *Compare, 41* (5), 615-628. doi:10.1080/03057925.2011.574969

Jones, L.L. & Golonka, S. (2012). Different influences on lexical priming for integrative, thematic, and taxonomic relations. *Frontiers in Human Neuroscience, 6,* Article 205, 1-17 (published 11 July 2012). doi:10.3389/fnhum.2012.00205

Kershner, R. (2009). Learning in inclusive classrooms. In P. Hick, R. Kershner & P. Farrell (eds) *Psychology for Inclusive Education: New directions in theory and practice* (pp. 52-65). Abingdon: Routledge.

Klein, J. T. (2014). Discourses of Transdisciplinarity: Looking back to the future. *Futures, 63C*, 68-74. doi:10.1016/j.futures.2014.08.008

Littleton, K. & Mercer, N. (2013) *Interthinking: Putting Talk to Work*. Abingdon: Routledge.

Moate, J. (2014). Dialogic struggles and pedagogic innovation. *Pedagogy, Culture and Society, 22* (2). 295-314. 10.1080/14681366.2014.891641

Ní Bhroin, O. (2013). Questions to assess learning as a communicative routine for inclusion. *British Journal of Special Education, 40* (3), 114-123. doi: 10.1111/1467-8578.12031

Noddings, N. (1992). *The Challenge to Care in Schools*, NY: Teachers College Press.

Norwich, B. (2014). Recognising value tensions that underlie problems in inclusive education. *Cambridge Journal of Education, 44* (4), 495-510. doi:10.1080/0305764X.2014.963027

Rajala, A., Hilppö, J., & Lipponen, L. (2012). The emergence of inclusive exploratory talk in primary students' peer interaction. *International Journal of Educational Research, 53*, 55-67. doi:10.1016/j.ijer.2011.12.011

Rietzchel, E.F., Nijstad, B.A. & Stroebe, W. (2007). Relative accessibility of domain knowledge and creativity: The effects of knowledge activation on the quantity and originality of generated ideas. *Journal of*

Experimental Social Psychology, 43, 933-946.
doi:10.1016/j.jesp.2006.10.014

Rose, J. & Norwich, B. (2014). Collective commitment and collective efficacy: a theoretical model for understanding the motivational dynamics of dilemma resolution in inter-professional work. *Cambridge Journal of Education, 44* (1), 59-74. doi:10.1080/0305764X.2013.855169

Salès-Wuillemin, E., Morlot, R., Fontaine, A., Pullin, W., Galand, C., Talon, D., & Minary-Dohen, P. (2011). Evoluation of nurses' social representations of hospital hygiene: From training to practice. *Revue européenne de psychologie appliquée, 61*, 51-63. doi:10.1016/j.erap.2010.06.001

Sheehy, K., & Rix, J.R.M. et al (2009). *A systematic review of whole class, subject-based, pedagogies with reported outcomes for the academic and social inclusion of pupils with special educational needs in mainstream classrooms.* In Research Evidence in Education Library, London: EPPI-Centre, Social Science Research Unit, Institute of Education.

Sheehy, K. (2013). Educational Psychology and the Development of Inclusive Education. *The Routledge International Companion to Educational Psychology,* 235-245. London: Routledge. Published online 19 July 2013.

Skidmore, D. (1999). Discourses of Learning Difficulty and the Conditions of School Improvement. *Educational Review, 51* (1), 17-28. doi:10.1080/0141192990250506

Smyth, F., Shevlin, M., Buchner, T., Biewer, G., Flynn, P., Latimier, C., Šiška, J., Toboso-Martín, M., Rodríguez Díaz, S., & Ferreira, M.A.V. (2014). Inclusive Education in Progress: Policy evolution in four European countries. *European Journal of Special Needs Education, 29* (4), 433-445. doi: 10.1080/08856257.2014.922797

Tharp, R.G. & Dalton, S.S. (2008). Orthodoxy, cultural compability, and universals in education. In J. Elliott & E. Grigorenko (Eds.) *Western Psychological and Educational Theory in Diverse Contexts* (pp.53-70). Abingdon: Routledge.

Tetler, S. & Baltzer, K. (2011). The climate of inclusive classrooms: the pupil perspective. *London Review of Education, 9* (3), 333-344. doi:10.1080/14748460.2011.616326

Thomas, G. & Loxley, A. (2001). *Deconstructing Special Education and Constructing Inclusion.* Buckingham: Open University Press.

Thomas, G. (2014). Epistemology and Special Education. In L. Florian (ed) *The SAGE Handbook of Special Education, 2nd Ed*. (pp. 267-279). London: SAGE.

Tragoulia, E. & Strogilos, V. (2013). Using dialogue as a means to promote collaborative and inclusive practices. *Educational Action Research, 21* (4) 485-505. doi:10.1080/09650792.2013.832342

Valls, R. & Kyriakides, L. (2013). The power of Interactive Groups: how diversity of adults volunteering in classrooms can promote inclusion and success of children of vulnerable minority ethnic population. *Cambridge Journal of Education, 43* (1) 17-33. doi:10.1080/0305764X.2012.749213

Wegerif, R. (2007). *Dialogic Education and Technology: Expanding the space of learning*. New York, NY: Springer.

Wegerif. R. (2008). Dialogic or Dialectic? The significance of ontological assumptions in research on educational dialogue. *British Educational Research Journal, 34* (3), 347-361. doi:10.1080/13540602.2011.554704

Ruth Kershner is Lecturer in Psychology of Education and Primary Education. University of Cambridge

Contact Address: Ruth Kershner. Faculty of Education University of Cambridge. 184 Hills Road, Cambridge, CB2 8PQ. United Kingdom. Email: rsk21@cam.ac.uk

Mindfulness as an Alternative for Supporting University Student Mental Health: Cognitive-Emotional and Depressive Self-Criticism Measures

Muhammad Abid Azam, Myriam Mongrain, Khushboo Vora, Meysam Pirbaglou, Saam Azargive, Tina Changoor, Noah Wayne, Crissa Guglietti, Alison Macpherson, Jane Irvine, Michael Rotondi, Dawn Smith, Daniel Perez, Paul Ritvo[1]

1) York University

PLEASE SCROLL DOWN FOR ARTICLE

Mindfulness as an Alternative for Supporting University Student Mental Health: Cognitive-Emotional and Depressive Self-Criticism Measures

Muhammad Abid Azam, Myriam Mongrain, Khushboo Vora, Meysam Pirbaglou, Saam Azargive, Tina Changoor, Noah Wayne, Crissa Guglietti, Alison Macpherson, Jane Irvine, Michael Rotondi, Dawn Smith, Daniel Perez, Paul Ritvo[1]
York University

Abstract

Increases in university-based mental health problems require alternative mental health programs, applicable to students with elevated psychological risks due to personality traits. This study examined the cognitive-emotional outcomes of a university mindfulness meditation (MM) program and their relationship with Self-Criticism (SC), a personality factor linked to depressive vulnerability. University students ($n = 71$) were assessed at baseline with the Depressive Experiences Questionnaire (DEQ), a measure of depressive personality traits, and two outcome measures: Profile of Mood States (POMS) and Anxiety Sensitivity Index (ASI). Students attending the MM program were reassessed for outcomes at 3 follow up assessments over 2 semesters. Repeated measures analyses of variance revealed improved within-subjects effects with large or very large effect sizes for the subsample that completed the MM program ($n = 18$) on the POMS Tension-Anxiety, POMS Depression, POMS Fatigue, and Anxiety Sensitivity Index. Multiple linear regression using self-criticism as a predictor of change in depressed mood revealed that higher self-criticism predicted greater reductions in POMS Depression. This study provides evidence for MM-related cognitive-emotional benefits, suggesting that students with elevated self-critical traits may derive exceptional benefits evident in greater reductions of depressed mood.

Keywords: mindfulness, meditation, mental health, university students, depression

Hipatia Press
www.hipatiapress.com

Mindfulness una Alternativa de Apoyo a la Salud Mental del Alumnado Universitario: Medidas Cognitivo-Emocionales y de Autocrítica Depresiva

Muhammad Abid Azam, Myriam Mongrain, Khushboo Vora, Meysam Pirbaglou, Saam Azargive, Tina Changoor, Noah Wayne, Crissa Guglietti, Alison Macpherson, Jane Irvine, Michael Rotondi, Dawn Smith, Daniel Perez, Paul Ritvo[1]
York University

Resumen

Este estudio examinó los resultados cognitivos y emocionales de un programa de meditación de atención plena (MAP) en una universidad en relación a la autocrítica, un factor de la personalidad vinculado con la vulnerabilidad depresiva. 71 estudiantes universitarios fueron evaluados al inicio del estudio con el Cuestionario de las Experiencias Depresivas, una medida de los rasgos de la personalidad depresiva, y dos medidas de resultado: el Perfil de los Estados de Ánimo (POMS) y el Índice de Sensibilidad a la Ansiedad. Los estudiantes que asistieron al programa MAP fueron reevaluados para los resultados en 3 seguimientos de evaluación durante 2 semestres. Medidas repetidas ANOVA reveló efectos mejorados intra-sujetos con grandes o muy grandes tamaños del efecto para la sub-muestra que completó el programa MAP (n = 18); POMS tensión-Ansiedad, Depresión POMS, POMS fatiga y Índice de sensibilidad ansiedad. La regresión lineal múltiple utilizando la autocrítica como predictor de cambio en el estado de ánimo deprimido reveló que una autocrítica mayor predijo mayores reducciones en POMS Depresión. Este estudio proporciona evidencia de beneficios cognitivos emocionales relacionadas con el MAP, e indica que los estudiantes con rasgos autocríticas elevados pueden derivar beneficios excepcionales evidentes en mayores reducciones de estado de ánimo deprimido.

Palabras clave: mindfulness, meditación, salud mental, universitarios, depresión

Hipatia Press
www.hipatiapress.com

A According to the National College Health Assessment, half of university students report above-average to high levels of psychological distress, with the assessed prevalence of anxiety and depressive disorders increasing from 20% to 24%, in just the last five years (ACHA, 2013). Student distress is exacerbated by inadequate student-counselor ratios (Lees & Davis, 2012), help-seeking barriers, and subclinical stress conditions, and mental health risk-elevating personality traits (e.g. maladaptive perfectionism) (Radhu, Daskalakis, Arpin-Cribbie, Irvine, & Ritvo, 2012) that can progress to diagnosable psychopathology if not preventively counteracted.

A longitudinal study of 16, 460 first-year undergraduate students in the United Kingdom showed steadily increasing depression and anxiety scores over the progressive course of their time in university academic programs (Cooke, Bewick, Barkham, Bradley, & Audin, 2006). Recent estimates indicate one-third of university students experience depressed-anxious moods (Eisenberg, Gollust, Golberstein, & Hefner, 2007) and one-quarter display maladaptive perfectionism (Radhu et al., 2012). The increasing rates and complexities of mental health disturbances have strained traditional university counseling resources (Lees & Davis, 2012), with cumulative data indicating a need for alternative mental health programs that would address the large numbers of students who fail to receive adequate support via traditional counselling. Furthermore, such programs could respond to recent data showing students prefer dealing with stress-related issues autonomously (73.3%) and believe they lack sufficient time for counseling treatment (46.7%) (Downs & Eisenberg, 2012)

Mindfulness Meditation and Mental Health Benefits

Scientific evidence is accumulating for the effectiveness of mindfulness meditation (MM) in treatment of anxiety-depressive disorders and reduction of distress. In MM research, individuals with self-critical tendencies are of particular interest, given their susceptibility to elevated stress (Grossman, Niemann, Schmidt, & Walach, 2004) which can be mitigated by addressing and reducing their experiences of symptomatic depression and anxiety (Hazlett-Stevens, 2012). One of the purposes of our research was to examine

the relationship between self-critical personality traits and improvements in psychological status in university students during MM program participation.

Mindfulness has been defined as increased awareness attained through deliberate, nonjudgmental attention to present time-frame experience (Didonna, 2009). During MM, participants attend to breathing sensations, and return attention to breathing after monitoring thoughts, emotions, and external stimuli. Attention to breathing sensations has several purposes: it orients one to a present time-frame, generates a relaxation effect, and can be readily resumed after distractive mind wandering (Grossman et al., 2004). With consistent practice, awareness of present-oriented experiences combined with a nonjudgmental attitude towards internal experiences, can reduce distress (Nyklícek & Kuijpers, 2008). Furthermore, current research suggests regular MM improves the neurophysiological health of cortical areas associated with attention and emotion regulation (Lazar et al., 2006).

Mindfulness and Depressive Self-Criticism

As MM has emerged as a potentially cost-effective mental health intervention, investigators have placed special focus on how it impacts depression. Psychometric data on trait features of mindfulness suggest protection against specific depressive symptoms (Grossman et al., 2004) while evidence from intervention research indicates implementation of Mindfulness-Based Cognitive Therapy (MBCT) helps reduce depressive symptoms and prevents relapse to diagnosable depressive disorder (Chiesa & Serretti, 2011). To date, minimal research has focused on how depressive personality vulnerabilities affect students undergoing MM training. To address this gap, our study focused on examining MM effectiveness while examining student's depressive vulnerabilities as indicated by a well-established personality questionnaire. This screening instrument, the Depressive Experiences Questionnaire (DEQ), is a measure of *depressive personality style* reflecting trait-like patterns predictive of subsequent incidence of diagnosable depression. The scale is comprised of three factors: Self-Criticism (defined as a failure in meeting standards, feelings of inferiority, and decreased self-worth), Dependency (reliance on interpersonal

relationships), and Efficacy (characterized by resilience in response to personal challenges) (Blatt, Zohar, Quinlan, Zuroff, & Mongrain, 1995). In previous validation efforts, a 12-month study found the Self-Criticism (SC) subscale predicted periods of depressive distress, particularly concerning autonomy and self-esteem (Viglione, Lovette, Gottlieb, & Friedberg, 1995).

The academic environment triggers self-critical states through frequent exposures to social-evaluative events. A past study revealed that negative automatic thoughts and anxiety sensitivity (i.e. fear of anxiety symptoms) mediated associations between symptoms of depression and anxiety (Pirbaglou et al., 2013). Theoretically, depressive self-criticism can be considered a trait manifestation of negative automatic thoughts, given the common characteristic of self-directed negative cognitive content. Self-critical personality vulnerabilities are, overall, highly relevant to student populations and therefore important to examine in better understanding the potential targets of preventive mental health interventions like MM.

Recently, a study explored the Big-Five personality traits and MM practice (van den Hurk et al., 2011) using a cross-sectional design. Experienced meditators and control subjects (meditators = 35, non-meditators; 35) were compared and those who meditated demonstrated positive associations with curiosity and positive affect, as well as negative associations with worry (van den Hurk et al., 2011). A randomized controlled trial recently found evidence for the effectiveness of an 8-week MM program in lowering stress in college students (Oman, Shapiro, Thoresen, Plante, & Flinders, 2008). Taken together, it seems the demanding and stress-inducing nature of the university environment is now well established, as are the mental health benefits of practicing MM. Accordingly, we were interested in implementing a low-cost, easily accessible MM training program within the university and evaluating its effectiveness.

The current paper extends prior research by examining the effectiveness of a specific MM training program designed for university students and exploring how students with self-critical traits. Our objectives were: i) to study changes in anxiety sensitivity and mood during the school year as students attended ongoing MM tutorials; ii) to study the relationship between self-critical traits and MM benefits over a two semester duration.

The paper reports on changes in mood and anxiety sensitivity before, during, and after an MM program and the predictive associations between Self Critical trait levels (at baseline) & depressed mood (at follow up). Participants were trained in MM over two continuous semesters, and two specific hypotheses guided evaluations:

(a) Participants will report progressive improvements in psychological well-being as measured by mood and anxiety sensitivity variables, reflecting MM effectiveness.
(b) Increased levels of Self-Criticism will predict greater reductions in depressed mood, reflecting the effectiveness of MM for those with markedly elevated Self-Critical personality traits. (This could also be understood via the corollary whereby lower levels of SC predict lesser reductions in depressed mood, following the premise that MM should have specific SC effects).

Methods

Participants

Seventy-one undergraduate students (43 females, 28 males) in a large public urban university participated in weekly MM sessions through the Fall and Winter semesters of the 2012/2013 academic year (Figure 1). Recruitment was facilitated through in-class announcements and poster-print advertisement. Online and paper surveys elicited self-report data reflecting mental health profiles at program initiation (baseline measures). Given the pragmatic study focus, the inclusion criteria were minimized: 1) currently enrolled in the university (part-time or full-time); and 2) computer access and/or smartphone access. Using a within-subjects design, participants were required to attend one MM tutorial per week during each semester and complete the psychological measures at three assessment periods; Time 2 (November- December), Time 3 (January-February), and Time 4 (March-April). To lessen participation demands on the students, follow-up assessments were provided in paper and electronic formats as needed. All

data were anonymized and maintained in a locked cabinet within a locked, secure office to ensure confidentiality. Study participants were given $50 cash payments for full study participation. This study met ethical guidelines and was approved by the Human Participants Review Subcommittee at York University where the study was conducted.

Intervention

The program progressively trained students in "mindfulness of breathing", a form of meditation frequently employed secularly (Didonna, 2009). Meditation tutorials were held on campus five times weekly, led by a faculty member and graduate students who were trained and experienced MM practitioners. The MM instruction standardly used guided participants to be seated on a chair or on the floor in a comfortable, attentive posture, free to close their eyes or keep them open. They were instructed to pay attention to breathing sensations and to mind-wandering events in a nonjudgmental manner. Tutorial leaders were provided ongoing supervision and training that combined investigations about their own MM experiences with explorations and feedback about the provision of verbal guidance to students during sessions. The tutorials were one hour in duration and typically involved 40-45 minutes of guided MM followed by a question-answer period based on participant experiences during and outside of the MM tutorials. Program participation required attendance at one tutorial per week, and participants were encouraged to conduct additional informal, independent MM practice.

Predictor Variable

Depressive Experiences Questionnaire (DEQ) (Viglione et al., 1995)**.** This 66-item questionnaire assesses levels of Self-Criticism and Dependency through a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). Dependency is characterized by submissiveness in interpersonal relationships and an intense desire for closeness coupled with fear of rejection while Self-Criticism, the measure emphasized in the current study, is defined by feelings of inferiority, guilt, and worthlessness, and

related to a sense of failure in meeting standards and expectations (Fuhr & Shean, 1992). Test-retest reliabilities for the DEQ factors are high, with test-retest correlations of .86 over a 3-4 week interval, .72 - .81 after a 13-week interval and of .79 for Dependency and Self- Criticism after a 1-year interval (Fuhr & Shean, 1992). Alpha coefficients have been previously calculated as .80 for Self-Criticism, and construct validity has been evidenced in a variety of contexts (Viglione et al., 1995). The DEQ was administered through the baseline survey, and standardized scores of self-criticism were calculated for each participant.

Outcome Variables

Profile of Mood States (POMS) (Bourgeois, LeUnes, & Meyers, 2010). The POMS is an adjective-based rating scale used to assess mood state and mood changes. It includes five negative mood subscales of Tension-Anxiety, Depression-Dejection, Fatigue-Inertia, Anger-Hostility, and Confusion-Bewilderment, and one positive mood subscale (Vigour-Activated). This study used a short 38-item version (Baker, Denniston, Zabora, Polland, & Dudley, 2002), and analyses focused on individual scores on the Tension-Anxiety (e.g., "Nervous"), Depression (e.g. "Discouraged") and Fatigue (e.g., "Worn Out") subscales. The alpha coefficients for Tension-Anxiety, Depression, and Fatigue have been found to be 0.79, 0.90, and 0.90 respectively (Bourgeois et al., 2010). Participants were asked to respond to the items based on how they felt at the time of questionnaire completion.

Anxiety Sensitivity Index (ASI) (Reiss, Peterson, Gursky, & McNally, 1986). The ASI is a 16-item inventory in which a 5-point Likert-based scale is used to rate the degree of concern about negative consequences of anxiety-related sensations. Anxiety sensitivity (AS) refers to a set of beliefs that physical anxiety symptoms have negative effects such as causing more severe distress, development of major illness, or socially detrimental consequences (Reiss et al., 1986). As the most widely validated established index of AS, the ASI has been found to have a factorial structure (Rodriguez, Bruce, Pagano, & Keller, 2004) that measures 3 factors: physical concerns ("It scares me when my heart beats rapidly"), mental incapacitation concerns ("When I can't keep my mind on a task, I worry that

I might be going crazy"), and fears of publicly observable anxiety reactions ("Other people notice when I feel shaky"). In this study, total ASI score was used to measure AS, as it represents the global-order anxiety sensitivity factor and therefore reflects cognitive-emotional processes related to anxious arousal (Rodriguez et al., 2004).

Statistical Analysis

We conducted repeated measures ANOVA to assess the effect of MM program over the study period (Table 2) in the subsample that completed all measures. These included pairwise tests between baseline measurements and scores obtained at the three subsequent time points. To capitalize on the relatively large sample available at Time 2 ($n = 37$), correlational analyses assessed the relationships between SC and POMS Depression in this subsample who underwent a full semester of MM training. The correlations between SC and POMS Depression were assessed at baseline, and SC and change scores were assessed at Time 2, with change scores calculated at each stage in reference to baseline measurements by subtracting scores at each time point from baseline scores. Change scores are referred to in the results section with the 'delta' symbol (Δ POMS Depression Time 2). Furthermore, as differing data were available at each assessment period due to participant attrition, we tested whether SC was associated with attrition using an independent samples t-test to explore SC effects. Finally, using multiple linear regression analyses, we assessed whether the significant relationships observed in the correlational analyses remained significant after controlling for potential confounding variables (Table 3). We used G*Power 3.1 (Erdfelder, Lang, & Buchner, 2007) to calculate post-hoc (observed) power for all ANOVA and regression models and SPSS statistical software (IBM Corp., 2011) for analytic procedures.

Results

Participation
For ethical, health and practical reasons, participants and interested students were permitted to attend MM tutorials regardless of participation in research

activities. Of the initial 71 students, 25.3% (n = 18) completed all four assessments (Figure 1). In addition, another 10 participants ceased completing questionnaires at various points but continued participation in meditation activities for the full duration of the study. Participants who dropped out and responded to inquiries most commonly cited workload or scheduling conflicts as reasons for ending participation, on a purely anecdotal basis.
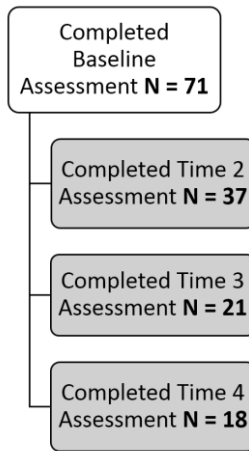

Figure 1: Completion of Assessments for Attending Mindfulness Meditation Program during the School Year. Flow chart of number of participants completing assessments throughout the MM program.

Participant attrition was a key study challenge which we believe is explained by multiple factors. The nature of the participant pool as individuals at elevated risk for anxiety, depression, and high sensitivity to critical thoughts (see baseline psychometric results) makes them more likely sensitive to academic pressures and university life stressors in ways than the general university student population. Particular evidence of their stress vulnerability was found in the high levels of drop-out surrounding the Fall examination period, as students studied for exams and completed final assignments critical to final grades. While the participants who remained

until the end of study were sufficient in number to conduct meaningful statistical analyses, the impact of attrition on the validity of these findings for those unable to persist with intervention and study assessments is acknowledged.

Table 1:
Baseline Demographics

Demographic Variable	Frequency	Percentage (%)*
Total Recruited	71	--
Age ($\mu = 25.5$; $\sigma = 10.17$)		
18-20	23	34.3
21-23	26	38.8
> 24	18	26.9
Missing	4	.056
Gender		
Female	43	66.2
Male	22	33.8
Missing	6	.085
Ethnicity		
West Asian	2	3.1
Black – African	5	7.7
Black/Indo – Caribbean	4	6.1
Caucasian	28	43.1
Latin American/Hispanic	2	3.1
Oriental (Chinese, Korean)	9	13.8
South Asian	10	6.5
South East Asian	5	7.7
Missing	6	.085

* Percentage value is calculated based on available data.

Only a minority of students had what might be considered substantial experience with mindfulness meditation, i.e. >50% of students had 0 to 10 hours of meditation experience, and a further 20% had accrued < 60 lifetime hours of meditation experience (Table 2). For context, a typical regular meditator (~30 minutes per day) would accumulate approximately 180 hours of practice per year.

Table 2.
Baseline Mental Health Characteristics

Characteristic	Frequency	Percentage (%)*
Hours of Meditation Experience		
None	19	28.4
< 10	34	50.8
10-60	13	19.4
> 60	1	1.5
Missing	4	.056
Depressive Disorder		
Yes, current	3	4.4
In the past	10	6.8
Never	55	80.9
Missing	3	.042
Anxiety Disorder		
Yes, current	5	7.4
In the past	9	13.2
Never	54	79.4
Missing	3	.042
Counseling/Psychotherapy		
Yes, currently	4	5.9
In the past	14	20.6
Never	50	73.5
Missing	68	.042
Psychiatric Medication		
Currently taking/Have taken	5	7.4
Never	63	92.6
Missing	3	.042

* Percentage value is calculated based on available data.

A total of 71 students completed the baseline measures; DEQ, POMS, ASI, and socio- demographic questionnaire (Table 1). The rate of self-reported anxiety and depression disorders (current or past) was 32%, with 12% reporting a current *clinical* diagnosis of an anxiety or depressive disorder. In order to better understand factors relating to dropout, chi-square tests of independence were performed to compare baseline statistics with dropouts (at Time 2) for demographic (gender, ethnicity) and all mental health characteristics. These tests showed that there were no significant

differences between Time 2 dropouts and non-dropouts on any demographic or mental health characteristics. Using the same dropout and non-dropout comparison, independent t-tests were conducted for psychometric scores (self-criticism, POMS tension, POMS depression, POMS fatigue, and anxiety sensitivity) at baseline, which did not result in significant differences, but showed higher mean scores on all the scales for the dropouts. This finding suggests that our speculation regarding student distress contributing to dropout risk merits further exploration and intervention, to be pursued in a future study.

Table 3:
One Way Repeated Measures ANOVA Results for Psychological Assessments of Student Meditators

Psychological Assessments (*n*)	$\bar{x}$ (σ)	p-value	Tests of Within Subject Effects Over Time	$\eta_p{}^2$
POMS Fatigue (18)			$F_{(3, 51)} = 45.70$	0.729
Time 1	15.72 (4.21)	--	(p < 0.001)	
Time 2	14.44 (4.63)	0.971		
Time 3	6.06** (4.30)	< 0.001		
Time 4	7.72** (4.51)	< 0.001		
POMS Tension-Anxiety (18)	15.72 (4.91)	--	$F_{(3, 51)} = 24.81$ (p < 0.001)	0.593
Time 1	15.72 (5.81)	1.000		
Time 2	5.56** (4.83)	< 0.001		
Time 3	7.61** (5.93)	0.001		
Time 4				
POMS Depression (18)	17.00 (6.77)	--	$F_{(3, 51)} = 17.93$ (p < 0.001)	0.514
Time 1	15.39 (8.12)	1.000		
Time 2	4.50** (7.12)	< 0.001		
Time 3	7.44* (7.84)	0.003		
Time 4				
ASI (18)			$F_{(3, 51)} = 69.01$	0.802
Time 1	32.94 (11.72)	--	(p < 0.001)	
Time 2	31.17 (10.96)	1.000		
Time 3	16.56** (12.05)	< 0.001		
Time 4	14.56** (11.38)	< 0.001		

* Significant at 0.01; ** Significant at 0.001 level (Adjustment for multiple comparisons: Bonferroni)
Notes: Time 2, Time 3, and Time 4 standard means were tested using Time 1 means in pairwise comparisons for each respective scale; $\bar{x}$ = standard mean; η_p^2 = partial eta squared; σ = Standard Deviation; ASI = Anxiety Sensitivity; POMS = Profile of Mood States

Table 3 displays results for one-way repeated measures ANOVAs conducted on each psychological measure to assess within-subject effects across time. All ANOVAs achieved an observed (post-hoc) power of greater than 80 percent (at the < 0.05 level) to detect a significant difference. Pairwise comparisons were conducted for psychological outcomes at three assessment periods compared to respective baseline scores, and were adjusted for multiple comparisons with the Bonferroni correction.

Supporting our hypotheses, significant within-subject effects indicated improvements for POMS Tension-Anxiety (F = 24.80, p < 0.001, η_p^2 = 0.73), POMS Depression (F = 17.93, p < 0.001, η_p^2 = 0.59), POMS Fatigue (F = 45.70, p < 0.001, η_p^2 = 0.51), and ASI (F = 69.01, p < 0.001, η_p^2 = 0.80). Pairwise comparisons did not reach significance at Time 2 indicating that by the November-December assessment, participants did not report substantial changes in mood and anxiety. However, at Time 3 and Time 4 (Figure 2) participants reported significant reductions in fatigue, tension, depression, and anxiety sensitivity. The effect sizes reflected in the η_p^2 estimates are all categorized as 'very strong', indicating the sample size was sufficient for the analyses undertaken (Cohen, 1988).

Correlational analyses examined the relationships between SC and POMS Depression Time 1 and Δ POMS Depression Time 2 (change in Depression scores at Time 2). There was a significant positive correlation between SC and POMS Depression (r = 0.46, p < 0.01), and a significant inverse correlation between SC and Δ POMS Depression Time 2 (r = -.32, p < 0.05). These associations indicated that improvements in depressed mood were predicted by SC. POMS Depression Δ Time 2 variable was verified as normally distributed prior to regression analysis.
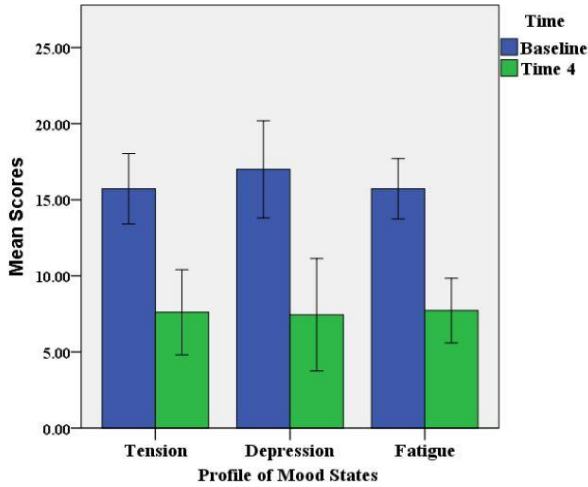
Figure 2. Changes in Profile of Mood States from Baseline to Time 4
$n = 18$; T-bars indicate standard error.

As correlational analyses were conducted on available data points at Time 2, it was important to rule out the possible confounding effect of dropout on SC associations. An independent samples t-test was used to compare self-criticism in the sub-sample who completed our MM program (completers) and all dropouts. There was no significant difference in SC scores for program completers ($\bar{x}$ = 0.116; SD = 0.154) vs. non-completers ($\bar{x}$ = 0.116; SD = 0.142); t(61) = -0.002, p = 0.999.

Table 4:
Regression Analyses for Self-Criticism as Predictor of Change in Depressed Mood (POMS)

Predictor	Unstandardized Coefficients	β	p-value
Self-Criticism	-3.19 (1.42)	-0.35	< 0.05
Dependency	-0.20 (1.44)	0.02	0.89
Age	-0.20 (0.11)	0.29	0.09

Notes: Outcome variable - POMS Depression Δ Time 2, Std. – Standard; β = standardized coefficient; Model: p < 0.05, r = 0. 46, r2 = 0.21, n = 37

Table 4 lists results of the regression model tested, using POMS Depression Δ at Time 2 as outcome and SC as primary predictor, with age and the Dependency subscale score as covariates. Dependency was included based on prior research strongly linking it with Self-Criticism (Mongrain & Leather, 2006). The model was significant overall (p < 0.05), and SC was found inversely related to reductions in depressed mood (i.e., higher Self-Criticism predicted greater reductions in Depressed Mood) at Time 2, accounting for 35% of variance (p < 0.05). It was noted that the sample (N = 37) and the analysis of 3 variables satisfied Cohen criteria of at least 10 subjects per variable (Cohen, 1988). However, the model achieved only modest observed (post-hoc) power of 58 percent (at the < 0.05 level) to detect a significant difference.

Discussion

This study found evidence for improvements in mood disturbance and anxiety sensitivity in university students through participation in an MM program. The significant within-subjects effects observed for psychological outcomes suggest gradual improvements in response to MM exposures with changes similar to those found in prior studies of MM and psychological status (Oman et al., 2008; Rosenkranz et al., 2013; Tang et al., 2007). Changes in the POMS subscales Tension-Anxiety, Depression, Fatigue, and overall ASI score were significant at Time 3 (Table 3), suggesting accumulated practice was associated with progressive, significant benefits. Although significant loss of participants to attrition occurred, participants were at elevated risk for anxiety and depression, and were highly sensitive to critical thoughts (see baseline psychometric results), making them more sensitive to academic pressures and university life stressors than the general student population. Particular evidence of this stress vulnerability was found in the high drop-out levels during the Fall exam period, as students studied for exams and completed assignments critical to final grades. While the participants who remained until the end of study were sufficient in number for valid, meaningful statistical analyses, the impact of attrition on the validity of these findings for those unable to persist (with intervention and study assessments) is acknowledged.

The use of change scores and the within-subjects design were aimed at mitigating the impact of between-subject variability. Acknowledging the need to improve retention in future research, these results are statistically sound and demonstrate meaningful improvements for participants in these domains. For these measures, scores remained improved from Time 3 to Time 4 indicating maintenance of improvements between the January-February and March-April measurements.

It should be noted that these improvements are particularly meaningful given the very high baseline levels of depression and distress reported by participants. At baseline, this group reported POMS subscale scores equivalent to or more distressed than depressed cancer patients (Baker, Denniston, Zabora, Polland, & Dudley, 2002) and adults seeking psychological treatment (Norcross, Guadagnoli, & Prochaska, 1984). Although our recruitment approach did not target individuals experiencing psychological distress, it is evident that students manifesting significant mood difficulties were drawn to the study, reiterating the alarming levels of mental health issues in universities (ACHA, 2013).Thus, it was encouraging to see that those who eventually completed the program demonstrated improvements. Results also supported the hypothesis that higher self-critical personality traits predicted greater improvements in depressed mood (Time 2).

Cognitive-Emotional Changes

Improvements in three negative subscales of the POMS were significant beginning at Time 3 and Time 4. These reductions in negative mood over time (51% for POMS Fatigue, 50% for POMS Tension, 74% for POMS Depression, and 56% for Anxiety Sensitivity) (Table 3) represent significant improvements, comparable to the effects found for alternative mental health programs in universities (Regehr, Glancy, & Pitts, 2013). These results compare favorably to prior mindfulness research using the POMS to study meditation-related mood changes in older adults undergoing an MBSR program (Young & Baime, 2010). Prior to MBSR, participants scored 7.4 on Tension-Anxiety, 5.4 on Depression, and 7.9 on Fatigue, and after MBSR, reported mean scores of 4.6, 3.3, and 4.9 respectively (Young & Baime,

2010). In the present study, participants started off with much higher baseline scores of Tension-Anxiety (15.72), Depression (17.0), and Fatigue (15.72), and afterwards reported means scores of 7.61, 7.44, and 7.72 respectively (Table 3). Training in MBSR might be considered significantly more resource and cost-intensive (seated postures, walking, yoga) than the one-hour per week commitment required for our MM program. The results of our study should be understood in the context of the more flexible requirements for meditation frequency (i.e. 1 hr. practice/week).

Over time, participants who continued mindfulness practice showed evidence of improved stress adaptation, with decreasing anxiety sensitivity (ASI) scores, indicating reduced fear of physical and psychological anxiety symptoms. This result builds on previous findings indicating that MBCT is associated with reductions in anxiety sensitivity over time (Chiesa & Serretti, 2011; Young & Baime, 2010). As can be readily recalled, our MM instructions focused on nonjudgmental attention to breathing sensations and mind-wandering events. With continued and supported MM practice, the reactivity to anxiety-related symptoms is likely to decrease over time, explaining this marked reduction in our participants.

Relationships between Self-Criticism and Depressed Mood

As hypothesized, higher baseline self-criticism scores predicted greater reductions in depressed mood at the second assessment (end of fall term). As mindfulness meditation practice specifically builds on the ability to self-regulate and gain unbiased perspectives, individuals who were less self-critical to begin with theoretically had less opportunity to benefit through improvements in that particular domain. Those with high levels of self-critical tendencies evidently experienced greater MM benefits through reductions in depressed mood. This finding is notable given the lack of research on the relationship between SC traits and MM benefits. Previous research has indicated that self-reported measures of breathing awareness during meditation and trait mindfulness (comprised of stable personality variables linked to mindfulness) were associated with the capacity to neutralize depressive cognitions (Frewen, Lundberg, MacKinley, & Wrath, 2011). The current finding between SC and the POMS Depression subscale

suggests that negative moods linked to depressogenic self-critical traits can be modified through MM practice.

Strengths & Limitations

This study used an informal, cost-effective, and pragmatic program to realistically assess MM benefits in a university setting, a departure from previously studied programs that were more costly and time-intensive and required more resources to implement. Our MM program was highly inclusive, efficient, and proved effective in a diverse academic population. Accordingly, this approach offers one potential model for future MM-based university programming and/or research.

　　While the present study did not include a control group, the effect sizes over multiple time period assessments were 'very strong,' indicating robust effects, following a within subjects design. The research demonstrating the prevalence of depression and anxiety in university students is abundant (Bayram & Bilgel, 2008; Downs & Eisenberg, 2012; Lynch, Gander, Kohls, Kudielka, & Walach, 2011; Sawatzky et al., 2012; Waller et al., 2005), and the group of participants in this study can reasonably be compared to those research findings. According to prior research, depression and anxiety tend to increase over the course of the academic year (Cooke et al., 2006), making our findings of significant reductions over time a significant reversal. As participation and regular attendance in the MM program was purely voluntary, the study was prone to selection bias. But significantly distressed students were drawn to the program (as evident in baseline measures) and those who persevered and remained with the program demonstrated significant benefits. As always, self-report questionnaire data may have been influenced by social desirability biases, general adaptations to the school year, and participant expectations related to MM benefits. Lastly, the small number of program completers due to a high dropout rate must be addressed in future iterations of a program of this kind so that more students are able to maintain attendance and derive MM benefits.

Conclusion

Our findings contribute to the growing MM literature, and to the emerging interest in disseminating mindfulness practices on university campuses for mental health benefits. In relating personality to the study of MM, the findings underline the importance of dispositional risk factors in mental illness and support the study of how mental health interventions can respond to risk-elevating traits such as self-criticism. This study also offers a model of implementing alternative, MM-based mental health programming for university students and depicts the realistic challenges involved with uptake and retention. Future investigations should explore other personality factors using a combination of psychological and neurophysiological variables to establish a more solid evidential foundation for alternative MM- based mental health programming on university campuses.

Competing Interests
The authors declare that they have no competing interests.

References

(ACHA), A. C. H. A. (2013). *Reference Group Executive Summary Spring 2013*. Hanover.

Baker, F., Denniston, M., Zabora, J., Polland, A., & Dudley, W. N. (2002). A POMS short form for cancer patients: Psychometric and structural evaluation. *Psycho-Oncology*, *11*(4), 273–281. doi:10.1002/pon.564

Bayram, N., & Bilgel, N. (2008). The prevalence and socio-demographic correlations of depression, anxiety and stress among a group of university students. *Social Psychiatry and Psychiatric Epidemiology*, *43*(8), 667–72. doi:10.1007/s00127-008-0345-x

Blatt, S. J., Zohar, a H., Quinlan, D. M., Zuroff, D. C., & Mongrain, M. (1995). Subscales within the dependency factor of the Depressive Experiences Questionnaire. *Journal of Personality Assessment*, *64*(2), 319–39. doi:10.1207/s15327752jpa6402_11

Bourgeois, A., LeUnes, A., & Meyers, M. (2010). Full-Scale and Short-Form of the Profile of Mood States : A Factor Analytic Comparison. *Journal of Sport Behavior*, *33*(4), 355.

Chiesa, A., & Serretti, A. (2011). Mindfulness based cognitive therapy for psychiatric disorders: a systematic review and meta-analysis. *Psychiatry Research*, *187*(3), 441–53. doi:10.1016/j.psychres.2010.08.011

Cohen, J. (1988). Statistical Power Analysis for the Behavioural Sciences. In *Statistical Power Analysis for the Behavioural Sciences* (Second Edi., pp. 273–407). Hillside, New Jersey: Lawrense Erlbaum Associates.

Cooke, R., Bewick, B. M., Barkham, M., Bradley, M., & Audin, K. (2006). Measuring, monitoring and managing the psychological well-being of first year university students. *British Journal of Guidance & Counselling*, *34*(4), 505–517. doi:10.1080/03069880600942624

Corp., I. (2011). IBM SPSS Statistics for Windows. Armock, NY: IBM Corp.

Didonna, F. (2009a). *Clinical Handbook of Mindfulness*. (F. Didonna, Ed.). New York, NY: Springer New York. doi:10.1007/978-0-387-09593-6

Didonna, F. (Ed.). (2009b). *Clinical Handbook of Mindfulness*. New York, NY: Springer New York. doi:10.1007/978-0-387-09593-6

Downs, M. F., & Eisenberg, D. (2012). Suicide and Suicide Prevention Suicidal College Students. *Journal of American College Health*, *60*(2).

Eisenberg, D., Gollust, S. E., Golberstein, E., & Hefner, J. L. (2007). Prevalence and correlates of depression, anxiety, and suicidality among university students. *The American Journal of Orthopsychiatry*, *77*(4), 534–42. doi:10.1037/0002-9432.77.4.534

Erdfelder, F., Lang, E., & Buchner, A.-G. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioural, and biomedical sciences. *Behaviour Research and Therapy*, *39*, 175–191.

Frewen, P., Lundberg, E., MacKinley, J., & Wrath, A. (2011). Assessment of Response to Mindfulness Meditation: Meditation Breath

Attention Scores in Association with Subjective Measures of State and Trait Mindfulness and Difficulty Letting Go of Depressive Cognition. *Mindfulness*, *2*(4), 254–269. doi:10.1007/s12671-011-0069-y

Fuhr, S., & Shean, G. (1992). Subtypes of Depression, Efficacy, and the Depressive Experiences Questionnaire. *The Journal of Psychology*, *5*(126), 495–506.

Grossman, P., Niemann, L., Schmidt, S., & Walach, H. (2004). Mindfulness-based stress reduction and health benefits. A meta-analysis. *Journal of Psychosomatic Research*, *57*(1), 35–43. doi:10.1016/S0022-3999(03)00573-7

Hazlett-Stevens, H. (2012). Mindfulness-based stress reduction for comorbid anxiety and depression: case report and clinical considerations. *The Journal of Nervous and Mental Disease*, *200*(11), 999–1003. doi:10.1097/NMD.0b013e3182718a61

Lazar, S. W., Kerr, C. E., Wasserman, R. H., Gray, J. R., Greve, D. N., Treadway, M. T., … & Fischl, B. (2005). Meditation experience is associated with increased cortical thickness.*Neuroreport, 16(17)*, 1893-1897.

Lees, J., & Davis, W. G. (2012). An Analysis of Counselling Services in Ontario Colleges Initial Report, (October).

Lynch, S., Gander, M.-L., Kohls, N., Kudielka, B., & Walach, H. (2011). Mindfulness-based Coping with University Life: A Non-randomized Wait-list-controlled Pilot Evaluation. *Stress and Health*, *27*(5), 365–375. doi:10.1002/smi.1382

Mongrain, M., & Leather, F. (2006). Immature Dependence and Self-Criticism Predict the Recurrence of Major Depression, *62*(6), 705–713. doi:10.1002/jclp

Norcross, J. C., Guadagnoli, E., & Prochaska, J. O. (1984). Factor Structure of the Profile of Mood States (POMS): Two partial replications. *Journal of Clinical Psychology*, *40*(5), 1270–1277.

Nyklícek, I., & Kuijpers, K. F. (2008). Effects of mindfulness-based stress reduction intervention on psychological well-being and quality of life: is increased mindfulness indeed the mechanism? *Annals of*

Behavioral Medicine : A Publication of the Society of Behavioral Medicine, *35*(3), 331–40. doi:10.1007/s12160-008-9030-2

Oman, D., Shapiro, S. L., Thoresen, C. E., Plante, T. G., & Flinders, T. (2008). Meditation Lowers Stress and Supports Forgiveness Among College Students : A Randomized Controlled Trial. *Journal of American College Health*, *56*(5). doi:10.3200/JACH.56.5.569-578

Pirbaglou, M., Cribbie, R., Irvine, J., Radhu, N., Vora, K., & Ritvo, P. (2013). Perfectionism, anxiety, and depressive distress: evidence for the mediating role of negative automatic thoughts and anxiety sensitivity. *Journal of American College Health : J of ACH*, *61*(8), 477–83. doi:10.1080/07448481.2013.833932

Radhu, N., Daskalakis, Z. J., Arpin-Cribbie, C. a, Irvine, J., & Ritvo, P. (2012). Evaluating a Web-based cognitive-behavioral therapy for maladaptive perfectionism in university students. *Journal of American College Health*, *60*(5), 357–66. doi:10.1080/07448481.2011.630703

Regehr, C., Glancy, D., & Pitts, A. (2013). Interventions to reduce stress in university students: a review and meta-analysis. *Journal of Affective Disorders*, *148*(1), 1–11. doi:10.1016/j.jad.2012.11.026

Reiss, S., Peterson, R. a, Gursky, D. M., & McNally, R. J. (1986). Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. *Behaviour Research and Therapy*, *24*(1), 1–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3947307

Rodriguez, B. F., Bruce, S. E., Pagano, M. A., & Keller, S. (2004). Factor Structure of the Anxiety Sensitivity Index in Longitudinal Study of Anxiety Disorder Patients. *Behaviour Research and Therapy*, *42*(1), 79–91.

Rosenkranz, M. a, Davidson, R. J., Maccoon, D. G., Sheridan, J. F., Kalin, N. H., & Lutz, A. (2013). A comparison of mindfulness-based stress reduction and an active control in modulation of neurogenic inflammation. *Brain, Behavior, and Immunity*, *27*(1), 174–84. doi:10.1016/j.bbi.2012.10.013

Sawatzky, R. G., Ratner, P. a, Richardson, C. G., Washburn, C., Sudmant, W., & Mirwaldt, P. (2012). Stress and depression in students: the

mediating role of stress management self-efficacy. *Nursing Research*, *61*(1), 13–21. doi:10.1097/NNR.0b013e31823b1440

Tang, Y.-Y., Ma, Y., Wang, J., Fan, Y., Feng, S., Lu, Q., … Posner, M. I. (2007). Short-term meditation training improves attention and self-regulation. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(43), 17152–6. doi:10.1073/pnas.0707678104

Van den Hurk, P. a M., Wingens, T., Giommi, F., Barendregt, H. P., Speckens, A. E. M., & van Schie, H. T. (2011). On the Relationship Between the Practice of Mindfulness Meditation and Personality-an Exploratory Analysis of the Mediating Role of Mindfulness Skills. *Mindfulness*, *2*(3), 194–200. doi:10.1007/s12671-011-0060-7

Viglione, D. J., Lovette, G. J., Gottlieb, R., & Friedberg, R. (1995). Depressive experiences questionnaire: an empirical exploration of the underlying theory. *Journal of Personality Assessment*, *65*(1), 91–9. doi:10.1207/s15327752jpa6501_7

Waller, R., Mahmood, T., Gandi, R., Delves, S., Humphrys, N., & Smith, D. (2005). Student mental health: how can psychiatrists better support the work of university medical centres and university counselling services? *British Journal of Guidance & Counselling*, *33*(1), 117–128. doi:10.1080/03069880412331335876

Young, L. a., & Baime, M. J. (2010). Mindfulness-Based Stress Reduction: Effect on Emotional Distress in Older Adults. *Complementary Health Practice Review*, *15*(2), 59–64. doi:10.1177/1533210110387687

Muhammad Abid Azam (PhD Candidate), **Myriam Mongrain** (PhD), **Khushboo Vora** (BSc), **Meysam Pirbaglou** (PhD Candidate), **Saam Azargive** (MSc), **Tina Changoor (PhD Candidate), Noah Wayne** (PhD), **Crissa Guglietti** (PhD), **Alison Macpherson** (PhD), **Jane Irvine** (PhD), **Michael Rotondi** (PhD), **Dawn Smith** (BSc), **Daniel Perez** (PhD Candidate), **Paul Ritvo** (PhD), at York University

Contact Address: Muhammad Abid Azam York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3

Three Scales to Measure Burnout of Primary School Teachers: Empirical Evidence on their Adequacy

Maria Platsidou, Athena Daniilidou[1]

1) University of Macedonia

PLEASE SCROLL DOWN FOR ARTICLE

Three Scales to Measure Burnout of Primary School Teachers: Empirical Evidence on their Adequacy

Maria Platsidou
University of Macedonia

Athena Daniilidou
University of Macedonia

Abstract

This study aimed at comparing the psychometric properties of three measures of burnout administered in 320 Greek primary school teachers, namely the Maslach Burnout Inventory (Maslach and Jackson 1982), the Burnout Measure (Pines and Aronson 1988) and the Copenhagen Burnout Inventory (Kristensen et al. 2005). Confirmatory factor analysis tested a three-factor solution for each inventory, consistent to the respective theory. Results yielded a reasonably good model fit for the MBI, and merely acceptable model fit for the BM and the CBI. Overall the findings suggest that the MBI is more appropriate instrument for assessing teachers' burnout compared to the MB and the CBI, which presented not so well-defined inner structure and highly correlated subscales.

Keywords: burnout syndrome, teachers, Maslach Burnout Inventory, Burnout Measure, Copenhagen Burnout Inventory

Hipatia Press
www.hipatiapress.com

Tres Escalas para Medir el Burnout del Profesorado de Primaria: Evidencias Empíricas de su Adecuación

Maria Platsidou
University of Macedonia

Athena Daniilidou
University of Macedonia

Resumen

El presente estudio se plantea con el objetivo de comparar las propiedades psicométricas de tres formas de medición del síndrome de *burnout* empleadas en 320 maestros de educación primaria griegos, concretamente, del Inventario del Síndrome de *Burnout* de Maslach (ISBM) (Maslach y Jackson, 1982), la Medición del Síndrome de *Burnout* (MSB) (Pines y Aronson, 1988) y el Inventario del Síndrome de *Burnout* de Copenhague (ISBC) (Kristensen *et al.,* 2005). Mediante la aplicación de un análisis factorial confirmatorio, se examinó una solución, planteada de manera consistente con la teoría a la que se refería, de tres factores para cada inventario. Los resultados obtenidos indicaron una adecuación del modelo razonablemente buena para el ISBM y simplemente aceptable para el MSB y el ISBC. En general, los hallazgos señalan que el ISBM constituye un instrumento más adecuado para la evaluación del *burnout* de los docentes en comparación con el MSB y el ISBC, ya que estos presentan una estructura interna bastante indefinida, además de contar con subescalas altamente relacionadas entre ellas.

Palabras clave:
síndrome de *burnout*, profesorado, inventario del síndrome de *Burnout* de Maslach, medición del síndrome de *burnout*, inventario del síndrome de *burnout* de Copenhague.

I n work-related settings, burnout refers to the exhaustion of employees' capacity to maintain an intense involvement that has a meaningful impact at work (Schaufeli, Leiter, & Maslach, 2009) and it impairs (profoundly or not) their emotional, mental and physical state and manifestations. Teaching profession is considered one of the most stressful professions due to the frequent and intense interaction with students, parents and peers (Hakanen, Bakker & Schaufeli, 2006). As a result, teachers are often driven to burnout. Ample research findings have shown that burnout of primary and secondary school teachers is high in most of the North European (Taris, Schaufeli, Schreurs & Calje, 2000), North American (Mearns & Cain, 2003), and Asian countries (Maslach, Schaufeli & Leiter, 2001), as well as in Australia (Kyriakou, 1987). Rather lower levels of burnout are reported by teachers in most of the Mediterranean countries such as Cyprus (Kokkinos, 2006), Israel (Pines, 2004), Turkey (Demirel, Güler, Toktamis, Özdemir, & Sezer, 2005) and Greece (Papastylianou, Kaila & Polychronopoulos, 2009; Platsidou &Agaliotis, 2008).Overall, research has not yet concluded whether differences in teachers' perceived burnout reflect varied levels of intense occupational stress or they imply that the meaning of burnout is not identical across countries and languages (Schaufeli et al., 2009).

Internationally, over 90% of the studies measuring teachers' burnout use the Maslach Burnout Measure (MBI, Maslach &Jackson,1986), although other relevant research instruments have also been proposed (Kristensen, Borritz, Villadsen & Christensen, 2005). Few studies have attempted to explore if preference of the MBI is warranted by its psychometric properties, comparing it to other scales, but they have not included recently designed scales. Therefore, this study was designed to compare measures of teachers' burnout obtained by three of the most widely used instruments, namely, the ones developed by Maslach and Jackson (1986), Pines and Aronson (1988), and Kristensen et al. (2005), respectively.

Maslach Burnout Measure (MBI)

The MBI is based on the Maslach and Jackson's (1986) model arguing that burnout arises as a result of prolonged work under pressure. The repeated failures to deal with tense and demanding situations in the workplace and the

decline of resistance to work stress may lead a person to burnout (Maslach & Schaufeli, 1993). According to this model, the syndrome of burnout is described by three dimensions which encompass different categories of symptoms: emotional exhaustion (feelings of being overextended and exhausted by work), depersonalization or cynicism (negative, uncaring attitudes toward recipients), and lack of personal accomplishment or reduced effectiveness (negative evaluation of performance and achievement in the job). Maslach and Jackson (1986) developed the MBI to assess the three aforementioned burnout dimensions of employees. Emotional exhaustion has been consistently viewed as the core, more stable and consistent component of the MBI (Schaufeli & Enzmann, 1998; Shirom, 2005), which usually appears first and predicts the other two components in longitudinal studies (Maslach & Schaufeli, 1993; Taris, Le Blanc, Schaufeli & Schreurs, 2005).

The MBI has been undoubtedly the most widely used instrument for measuring burnout in a variety of occupations, cultures and work settings (Schutte, Toppinen, Kalimo, & Schaufeli, 2000).Over the many years of research, ample evidence has been provided in support that the MBI is a reliable and valid measure of teacher burnout. Specifically, many factor analytic studies of teachers have reported three well-defined burnout factors representing emotional exhaustion, depersonalization, and reduced personal accomplishment with good internal consistency and test-retest reliability, as well as satisfactory convergent and discriminant validity (e.g., Demerouti, Bakker, Vardakou, & Kantas, 2003; Kokkinos, 2006; Papastylianou et al., 2009; Kantas & Vassilaki, 1997).Another asset of the MBI is that cut-off points are determined statistically, for instance, for "low", "average", and "high" scores of burnout, thus making test results easier to interpret (Maslach et al., 1996).

On the other hand, the MBI is criticized for being developed inductively and not stemming from a sound theoretical model (Shirom, 2005). As Kristensen et al. (2005, p. 193) note, the MBI and the Maslach definition of burnout "have become two sides of the same coin: Burnout is what the MBI measures, and the MBI measures what burnout is". In addition, they argue that depersonalisation is a coping strategy developed in a specific situation rather than a component of the burnout syndrome. As far as personal

accomplishment is concerned, some studies have found that it develops largely independent from the other two burnout dimensions (Schutte et al., 2000); therefore, they claim it may not be part of the total concept of burnout. Moreover, Kristensen et al. (2005) critically note that the Maslach definition of burnout is applied exclusively to employees in the human service sector.

In prior studies, internal consistency of the MBI ranges from .70 to .90 (Maslach et al., 1996). Studies conducted in Greece and Cyprus have pointed satisfactory reliability indices for emotional exhaustion (α = .82 to .85) and personal accomplishment (α = .72 to .75) and low to adequate reliability indices for depersonalization (α =53 to .63) (Kokkinos, 2006; Platsidou & Agaliotis, 2008).

Burnout Measure (BM)

Pines and her colleagues defined burnout as a state of physical, emotional and mental exhaustion caused by long term exposure to emotionally demanding situations (Pines & Aronson, 1988). "*Physical exhaustion* is characterized by low energy, chronic fatigue and weakness. *Emotional exhaustion*, the second component of burnout, involves primarily feelings of helplessness, hopelessness and entrapment. *Mental exhaustion*, the third component, is characterized by the development of negative attitudes towards one's self, work and life itself" (Pines & Aronson, 1988, p.12). Tedium was proposed as an alternative term to describe burnout (Pines, Aronson & Kafry, 1981). According to Pines, burnout is caused by the existential need of persons to give their life a meaning and, therefore, it is applicable to all life domains such as marital relationships (Pines, 1996) and aftermath of political conflicts (Pines, 1994). In the context of employment, in particular, Pines (2002) claims that when work fails to meet the person's need for a meaningful life, burnout is inevitable. She further emphasizes the role of work environment, while personality traits and motives contribute only in triggering and determining the severity of the syndrome. The most dedicated employees exhibit more severe forms of burnout.

Pines et al. (1981) developed the Burnout Measure (BM) scale to assess physical, mental and emotional exhaustion of people at work, which later

became a very popular measure of burnout. Despite their multidimensional conception of burnout, Pines and Aronson (1988) do not place the three types of exhaustion into a theoretical framework and they do not provide explanations on how they develop, differentiate or interact over time (Enzmann, Schaufeli, Janssen, & Rozeman, 1998). Consequently, although they classify the BM items into the three types of exhaustion described earlier, they compute one single burnout score. Subsequent research has produced ambivalent findings regarding the dimensionality of the BM: some studies have succeeded (Weisberg & Sagie, 1999) while others have failed (Ray & Miller, 1991; Schaufeli & Van Dierendock, 1993) to support the hypothesized three-factor solution. At the same time, many researchers use the BM as a unidimensional or global construct (Corcoran, 1986; Pines & Aronson, 1988). Finally, other studies have proposed a different pattern of dimensionality of the BM; for example, Schaufeli and van Dierendonck (1993) identified three correlated factors representing demoralization, exhaustion, and loss of motive; Shirom and Ezrachi (2003) argued that the BM structure is clearly multidimensional with imperfectly correlated constituent components. In relevant studies, internal consistency of the BM subscales ranged from .80 to .90 (Pines & Aronson, 1981, 1988).

Copenhagen Burnout Inventory (CBI)

The Copenhagen Burnout Inventory (CBI) was proposed by Kristensen and his colleagues (2005) as a response to the harsh criticism leveled by themselves and other researchers to the model and, especially, the research instrument of Maslach and her partners. According to the theoretical considerations of the CBI, burnout may be conceptualized broadly as a state of exhaustion which is both general (personal burnout) and specific (work and client burnout) (Borritz & Kristensen, 1999a, 1999b). Specifically, this research instrument assesses burnout in three subscales (Kristensen et al., 2005): *personal burnout* (the experience of physical and psychological fatigue and exhaustion), *work-related burnout* (refers to the person's own attribution of burnout symptoms to their work) and *client-related burnout* (the extent to which people attribute their exhaustion and fatigue to factors related to their ''people work''). Personal burnout, the generic part of the

CBI, may also occur among those who do not work such as young people, unemployed, early retired people, pensioners, and housewives. In the other two subscales, questions are formulated in such a way that they can be answered by employees of all professions. As Kristensen et al. (2005, p. 205) claim, "the three scales can be used in different domains (all persons, persons who work, and persons who do client-work)". Consequently, "in many concrete studies it would be meaningful to use only one or two of the scales".

As a recently developed scale, the CBI has still a long way to go to prove it is a valid and reliable scale for measuring burnout. So far, it has been translated into many languages and is being used in many cultures such as Japan (Odagiri, Shimomitsu, Ohya, & Kristensen, 2004) and China (Lin & Lin, 2013). Preliminary findings have shown that the CBI presents very good psychometric properties (e.g., high internal reliability, high convergent and divergent validity) in Danish (Kristensen et al., 2005) and Australian samples in various professions (Winnwood & Winefield, 2004), as well as in paper-and-pencil and online administration (Campos, Zucoloto, Bonafe, Jordani & Maroco, 2011). Its internal consistency and homogeneity, factorial validity and criterion-related validity were also found to be acceptable in a study of teachers in New Zealand (Milfont, Denny, Ameratunga, Robinson, & Merry, 2008). Yet, many more studies having used the CBI in the last years do not report its psychometric properties (D' Souza, Egan & Rees, 2011; Klein, Grosse Frie, Blum, & Knesebeck, 2010).

Aims and Hypotheses of the Present Study

In this context, the study aimed at comparing three commonly used measures of teachers' burnout (MBI, BM and CBI), which are based on the three alternative theoretical models described earlier. More specifically, our first aim was to investigate the factorial validity and reliability of the three instruments administered in teachers. According to our hypothesis, for each of the three burnout inventories, a correlated three-factor model consistent to the original inventories will fit the data. Secondly, we investigated the interrelations of the different burnout measures to test for convergent validity. To support convergent validity, correlations between the MBI-

emotional exhaustion scale, the BM and the CBI subscales are expected to be medium to high, since they all address a feeling of overtiredness focused on various domains (emotional, physical, mental) or sources (person, work, client). Depersonalization expresses a different aspect of burnout, so correlations with the above subscales are expected to be significant but lower and positive. Finally, reduced personal accomplishment indicates profound and all-out burnout, thus, it is expected to correlate moderately with the other burnout subscales.

Method

Participants

In this study, 320 primary school teachers working in public schools were tested, excluding teachers of specific teaching subjects such as music, sports, foreign languages etc. Data were collected from 49 state schools in urban and suburban areas of Greece. Of the participant teachers, 200 were female (62.5%) and 120 male (37.5%). The age of the sample ranged between 23 and 57 years, with a mean of 42.03 years ($SD = 8.88$). In relation to their marital status, 217 were married or in a permanent relationship (67.8%), 101 were single (31.6%) and 2 were widowed (0.6%). As to their job characteristics, total teaching experience ranged from 1 to 34 years, with a mean of 16.23 years ($SD = 8.53$). One hundred and six teachers (33.1%) had an MA or a PhD degree, while 214 (66.9%) held a BA degree in Primary Education.

Research Instruments

Teachers' burnout was assessed using three of the most commonly used inventories of burnout. The first one (MBI) has been the measure of choice in most relevant studies of Greek teachers (e.g., Kantas & Vassilaki, 1997; Papastylianou et al., 2009; Platsidou & Agaliotis, 2008). The other two inventories had not been used before in the Greek context, to the best of our knowledge; so translation into Greek was done first with the help of a

professor of English language and following the standard procedure for translation.

Maslach Burnout Inventory (MBI)

The Greek version of the Maslach Burnout Inventory (Maslach et al., 1996, translated by Kokkinos, 2006) was administered to the teachers. It contains 22 items that fall into three subscales: emotional exhaustion (9 items), depersonalization (5 items) and lack of a sense of personal accomplishment (8 items). Participants rated how frequently they experience these feelings on a 7-point scale, ranging from 0 (*Never*) to 6 (*Daily*).

Burnout Measure (BM)

The Burnout Measure was developed by Pines and Aronson (1988) to measure the three dimensions of burnout suggested by their theory: emotional exhaustion (7 items), physical exhaustion (7 items) and mental exhaustion (7 items). Teachers were asked to assess how frequently they experience the feelings described in the 21 items during the last 4-6 weeks using a 7-point scale, ranging from 0 (*Never*) to 6 (*Daily*). Scoring was reversed when necessary so that higher scores indicate elevated levels of burnout.

Copenhagen Burnout Inventory (CBI)

This inventory designed by Kristensen et al. (2005) consists of 19 items measuring personal burnout (6 items), work-related burnout (7 items) and client-related burnout (6 items). Participants rated how frequently they experience these feelings on a 5-point scale, ranging from 0 (*Never*) to 4 (*Always*). Scoring was reversed where necessary so that higher scores indicate higher levels of burnout.

Procedure

Teachers were contacted either in person at their schools or via email by the second author and were asked to participate in the study voluntarily with no other incentive being offered. They were assured that their responses would be treated with confidentiality and strictly for the research purpose, and they would not be provided to superiors. After obtaining their informed consent to participating, the inventories were administered for completion. The usual time of completion was approximately 15 min. Administration of the inventories took place from February to April 2014, that is, at the second half of the school year.

Data Analysis

Initially, a series of confirmatory factor analyses (using the maximum likelihood robust method of estimation) was performed to test the factorial validity of the three burnout instruments; analyses were run with the EQSWIN 6.1 program. To assess model fit we used well-established indices such as the x2-to-df ratio, CFI, SRMR, and RMSEA (Kline, 2005).In addition, reliability coefficients and means of the derived factors were computed. To examine convergent validity, correlations between the different burnout subscales were estimated.

Results

Factorial Validity of the Three Burnout Instruments

A correlated three-factor model was tested for each instrument in which the items loaded on their respective latent factor and the three latent factors were allowed to correlate. As Table 1 illustrates, factorial validity of the MBI was confirmed. Specifically, the ratio χ^2 /df = 1.82, CFI = .918, SRMR = .065 and RMSEA=.051 (CI90% .042–.059) indicated a reasonably good model fit. As expected, correlations between burnout dimensions were at medium to low levels and at the predicted direction. Reliability of the three subscales

was satisfactory (α =.87 for emotional exhaustion, α = .69 for depersonalization and α = .88 for personal accomplishment).

Table 1

The structure of the Maslach Burnout Inventory (standardized solution) in the test sample

Factors					
Items	ME	MP	MD	E	R^2
ME1	.678			.735	.459
ME2	.734			.679	.539
ME3	.715			.699	512
ME4	.849			.528	.721
ME5	.712			.703	.506
ME6	.357			.934	.127
ME7	.673			.739	.454
ME8	.762			.648	.580
ME9	.596			.803	.355
MP10		.529		.848	.280
MP11		.741		.672	.549
MP12		.717		.698	.513
MP13		.704		.710	.495
MP14		.783		.622	.614
MP15		.709		.705	.503
MP16		.689		.725	.474
MP17		.711		.703	.505
MD18			.403	.915	.162
MD19			.845	.535	.714
MD20			.745	.667	.555
MD21			.426	.905	.181
MD22			.348	.938	.121
Factor Correlations					
F1 – F2	-.206				
F2 – F3	-.388				
F1 – F3	.385				

Note: ME = Emotional Exhaustion; MP = Personal Accomplishment; MD = Depersonalization. All factors loadings, correlation and covariance indices are significant (p < .05).

The hypothesized factorial validity of the BM was also confirmed (see Table 2) although, in this case, statistical indices demonstrated a barely acceptable model fit (Kline, 2005): χ^2 /df = 3.6, CFI = .902, SRMR = .075 and RMSEA=.079 (CI90%   .071–.086). The three burnout factors were moderately to highly correlated, especially the emotional with the mental exhaustion. Reliability indices of the BM subscales were very good: α = .88 for emotional exhaustion, α = .83 for mental exhaustion, and α = .92 for physical exhaustion.

Table 2

The structure of the Burnout Measure (standardized solution) in the test sample

Factors					
Items	PE	PM	PP	E	R^2
PE1	.700			.714	.490
PE2	.816			.578	.666
PE3	.781			.625	.610
PE4	.569			.823	.323
PE5	.671			.741	.451
PE6	.752			.660	.565
PE7	.663			.748	.440
PM8		.708		.706	.502
PM9		.634		.773	.402
PM10		.706		.708	.499
PM11		.624		.781	.389
PM12		.708		.707	.501
PM13		.528		.849	.279
PM14		.565		.825	.319
PP15			.796	.605	.634
PP16			.868	.496	.754
PP17			.886	.463	.785
PP18			.737	.676	.543
PP19			.867	.499	.751
PP20			.866	.500	.750
PP21			.369	.929	.136

Factor Correlations	
F1 – F2	.865
F2 – F3	.512
F1 – F3	.617

Note: PE = Emotional Exhaustion; PM = Mental Exhaustion; PP = Physical Exhaustion. All factors loadings and correlation and covariance indices are significant ($p < .05$).

Similar to the above, the fit of the model testing factorial validity of the CBI was mediocre (Kline, 2005): $\chi^2/df = 2.92$, CFI = .898, SRMR = .062 and RMSEA =.078 (CI90% .069–.086). Taken these reservations into consideration, we can tentatively assume that the hypothesized internal structure of the CBI was confirmed, as Table 3 presents. However, the three factors were highly or very highly correlated. Cronbach alphas indicated good reliability for the three CBI subscales ($\alpha = .90$ for personal burnout, $\alpha = .84$ for work burnout and $\alpha = .78$ for client burnout).

Table 3
The structure of the Copenhagen Burnout Inventory (standardized solution) in the test sample

Factors					
Items	CP	CW	CS	E	R^2
CP1	.798			.603	.636
CP2	.807			.591	.651
CP3	.820			.572	.673
CP4	.757			.653	.573
CP5	.835			.550	.698
CP6	.651			.759	.423
CW7		.686		.728	.470
CW8		.843		.538	.710
CW9		.590		.807	.348
CW10		.753		.658	.566

(continued)

Factors				
CW11	.511		.860	.261
CW12	.618		.786	.382
CW13	.527		.850	.278
CS14		.661	.750	.437
CS15		.696	.718	.485
CS16		.609	.793	.371
CS17		.479	.878	.229
CS18		.660	.751	.436
CS19		.580	.815	.336
Factor Correlations				
F1 – F2	.904			
F2 – F3	.796			
F1 – F3	.706			

Note: CP = Personal Burnout; CW = Work-Related Burnout; CS = Student-Related Burnout. All factors loadings and correlation and covariance indices are significant ($p < .05$).

Correlations among the Three Burnout Inventories

Subsequently, correlation matrix of the different burnout measures was obtained. Inspection of Table 4 largely confirms our hypothesis. In general, the MBI-emotional exhaustion and the BM and the CBI subscales were moderately to highly correlated. Depersonalization had low bar significant correlations with the other burnout measures, but its correlation with the BM-physical exhaustion was not significant. Finally, personal accomplishment correlated negatively and moderately low with the BM and the MBI measures, with all correlations being significant. Overall, the mean of the CBI subscales presented higher correlations with both the mean of the MBI ($r = .683$, p < .001) and the mean of the BM subscales ($r = .706$, p < .001) than the correlations between the last two ($r = .545$, p < .001). The differences between these correlations were found significant, $Z = -4.32$, *p*< .001 and $Z = -3.34$, *p*< .001, respectively).

Table 4:
Correlation matrix of the burnout subscales.

	1	2	3	4	5	6	7	8
MBI								
1. Emot. exhaust.								
2. Pers. accompl.	-.197**							
3. Depersonaliz.	.310**	-.388**						
BM								
4. Emot. exhaust.	.552**	-.235**	.215**					
5. Mental exhaust.	.515**	-.291**	.227**	.711**				
6. Physical exhaust.	.610**	-.143*..	.095..	.638**	.574**			
CBI								
7. Personal burnout	.684**	-.256**	.191**	.686**	.567**	.748**		
8. Work burnout	.728**	-.363**	.273**	.562**	.547**	.627**	.776**	
9. Client burnout	.559**	-.360**	.320**	.409**	.358**	.390**	.587**	.728**

*Note: * $p < 0.5$, ** $p < .001$*

Discussion

In this study, three commonly used measures of burnout were administered to Greek primary school teachers with the intention of comparing their psychometric properties. It is noted that the MBI has been widely used in Greek samples so far, but the BM and the CBI were used for the first time, so their factor structure was tested first.

Factorial Validity of the Three Burnout Inventories

Confirmatory factor analysis of the hypothesized three-factor structure yielded a reasonably good model fit for the MBI, and merely acceptable model fit for the BM and the CBI. More specifically, the Greek version of the MBI was found to assess the three dimensions of teachers' burnout as predicted by the theory and their internal structure matched the original MBI (Maslach & Jackson, 1986; Maslach et al., 1996). Reliability of the three subscales was satisfactory. This finding is consistently confirmed in other studies; in particular, invariant factorial validity and good reliability of the MBI have been noticed in most of the international (Boles et al., 2000; Byrne, 1991) and the Greek studies of teachers (Papastylianou et al., 2009; Platsidou & Agaliotis, 2008), or other professionals (e.g., managers, clerks, foremen technicians, Schutte et al., 2000). It is interesting to note that factorial structure of the MBI, as well as the CBI, was invariant between paper-and-pencil and online administration and reliability was good in both formats (Campos et al., 2011). Against this backdrop, it can be easily comprehended why MBI has become so popular among researchers in so many countries, in spite of the criticism is has received (Kristensen et al., 2005; Shirom, 2005).

For the BM, confirmatory factor analysis revealed that the model of three subscales proposed by Pines and Aronson (1988, physical, emotional and mental exhaustion) adequately fits the data and the internal consistency of the three subscales was very good. Prior findings are inconclusive regarding the latent structure of the BM. Some studies (as reviewed by Enzmann et al., 1998), considered the BM a unidimensional scale, while in other studies a three-factor structure was identified, yet not corresponding to the dimensionality originally proposed. Although merely acceptable, the model fit of the present study confirmed the three subscales of the original BM. This alone is an important finding, given that no prior study has confirmed the proposed model in teacher samples. However, the high correlations between the three latent factors indicate the possibility of a general, higher order factor; this may support the argument that the BM captures a particular aspect of burnout related to fatigue and tiredness (Enzmann et al., 1998; Shirom & Ezrachi, 2003).

Finally, the factorial structure of the CBI was tested. Kristensen et al. (2005) provided empirical support of the reliability and validity of the CBI and showed it was adapted to different professional occupations, but they argued it should also be tested in different cultures. In the present study, the CBI is administered for the first time in Greek teachers. A three-factor model was tested, in which the CBI items load on their respective latent factor (personal, work and client burnout). This model was found to have a mediocre fit to the data. Internal consistency of the three latent factors was very good, however they were found to be highly correlated. High correlations between these factors were also found in prior studies, thus bringing into question whether the CBI subscales possess adequate discriminant validity, whether a general, higher order burnout factor is possible, or whether calculation of one single burnout score from the CBI is more justified rather than the three subscales(Milfont et al., 2008; Winwood & Winefield, 2004).

Relations among Burnout Measures

In the next step, correlations among all burnout measures were estimated. The MBI-emotional exhaustion scale, the BM and the CBI subscales were found to be medium to highly intercorrelated. This finding supports convergent validity of the above burnout measures as it confirms that they all address a feeling of emotional, physical, or mental overtiredness related to the person, work, or client domain. As predicted, reduced personal accomplishment and depersonalization had, respectively, moderately low and low correlations with the BM and the CBI subscales. Correlations were significant and to the predicted direction, except for the correlation of depersonalization with physical exhaustion which was not significant.

Based on the above, it can be assumed that the MBI subscales actually assess different aspects of burnout, as described by the theoretical model (Maslach & Jackson, 1986). On the other hand, the BM and the CBI subscales seem to assess quite similar manifestations or experiences of burnout and therefore the extraction of a single score for each inventory would be more justified rather than having a three-subscale solution. In agreement to the above, prior studies have reported that the discriminant

validity of the BM is weak and concluded that the BM is not suitable for the measurement of burnout as a distinct phenomenon or for differentiating burnout from the related but distinct affective states of anxiety. Instead, they suggest it may be adequate as a general measure of psychological distress (Enzmann et al., 1998; Shirom & Ezrachi, 2003; Schaufeli & Van Dierendonck, 1993).

Likewise, earlier findings regarding the CBI have shown considerable dependency between the three burnout scales as they were highly intercorrelated (Milfont et al., 2008). Kristensen et al., (2005) notice that correlations between the three CBI scales varied considerably across different workplaces and, in general, are lower in other professions. Apparently, findings from the present study confirm that these correlations are fairly high in the teaching profession.

In conclusion, combined with the results of confirmation factor analysis discussed earlier, it appears that the MBI is more appropriate instrument for assessing teachers' burnout compared to the BM and the CBI, which presented not so well-defined inner structure and highly correlated subscales. On the other hand, earlier research has shown that the MBI and the CBI indicated substantial similarity in the overall proportion of respondents identified as manifesting high burnout (Winwood & Winefield, 2004), which implies that the two measures of burnout converge. Evidently, more research is needed in the future to test the adequacy of the three-subscales of the BM and the CBI to fit data obtained by teachers in Greece as well as in other countries; many more findings need to be accumulated before the MBI is dethroned as the most preferred inventory for measuring teachers' burnout.

Conclusions and Limitations

In a nutshell, the present study confirmed the factorial structure of the MBI as proposed by its designers in our sample of Greek primary school teachers. In parallel, factorial structure of the BM and the CBI was also confirmed, although these models fitted adequately to the data. Furthermore, it showed that the MBI assesses different aspects of the burnout phenomenon, while in the other two measures subscale scores were fairly intercorrelated indicating that they might assess burnout as a more unified phenomenon. Based on the

above, one could be advised to use the MBI if she wants to check differentiated aspects of teachers' burnout, or prefer the BM or the CBI if she seek to obtained a more unified measurement of burnout and with no need to reverse any scores.

The present study had certain limitations. Specifically, results do not apply to all teachers in the educational hierarchy and are likely to have been changing as a result of the financial situation of the country and the ongoing educational reforms that are causing stress and insecurity to all employees. Moreover, a measure of occupational stress could be used as a criterion variable to get a more comprehensive testing of the discriminant validity of the burnout measures in relation to familiar stress reactions.

References

Boles, J. S., Dean, D. H., Ricks, J. M., Short, J. C., & Wang G. (2000). The dimensionality of the Maslach Burnout Inventory across small business owners and educators. *Journal of Vocational Behavior*, 56, 12-34. doi:10.1177/1094428104268027

Borritz, M., & Kristensen, T. (1999a). *Copenhagen Burnout Inventory* (1st Ed.). Copenhagen, Denmark: National Institute of Occupational Health.

Borritz, M., & Kristensen, T. (1999b). *PUMA (Study on Personal Burnout, Work Burnout and Client Burnout)*. Copenhagen, Denmark: National Institute of Occupational Health.

Byrne, B. M. (1991). Burnout: Investigating the impact of background variables for elementary, intermediate, secondary, and university educators. *Teaching and Teacher Education*, 7, 197–209. doi: 10.1007/s11205-009-9533-7

Campos, J. A. D. B., Zucoloto, M. L., Bonafé, F. S. S., Jordani, P. C., & Marôco, J. (2011). Reliability and validity of self reported burnout in college students: A cross randomized comparison of pencil-and-paper vs. online administration. *Computers in Human Behavior, 27*(5), 1875-1883. doi: 10.1590/S1807-25772013000500002

Corcoran S. M. (1986). Planning by expert and novice nurses in cases of varying complexity. *Research in Nursing and Health*, 9, 155-162. doi: 10.1002/nur.4770090211

Demerouti, E., Bakker, A. B., Vardakou, I., & Kantas, A. (2003). The convergent validity of two burnout Instruments: A multitrait - multimethod analysis. *European Journal of Psychological Assessment*, *18*, 296-307. doi: 10.1027//1015-5759.19.1.12

Demirel, Y., Guler, N., Toktamis, A., Ozdemir, D. & Sezer, R.E. (2005) Burnout among High School Teachers in Turkey. *Middle East Journal of Family Medicine*, *3*(3), 33-36.

D'Souza, F., Egan, S. J., & Rees, C. S. (2011). The relationship between perfectionism, stress and burnout in clinical psychologists. *Behavior Change*, *28*, 17-28. doi: 10.1375/bech.28.1.17

Enzmann, D., Schaufeli, W. B., Janssen, P., & Rozeman, A. (1998). Dimensionality and validity of the Burnout Measure. *Journal of Occupational and Organizational Psychology*, *71*, 331-351. doi: 10.1111/j.2044-8325.1998.tb00680.x

Hakanen, J., Bakker, A.B., & Schaufeli, W.B. (2006). Burnout and work engagement among teachers. *The Journal of School Psychology*, *43*, 495-513. doi:10.1016/j.jsp.2005.11.001

Kantas, A., &Vassilaki, E. (1997). Burnout in Greek teachers: Main findings and validity of the Maslach Burnout Inventory. *Work and Stress*, *11*, 94-100. doi:10.1080/02678379708256826

Klein, J., Grosse Frie, K., Blum, K., & Knesebeck, O. v. d. (2010) Burnout and perceived quality of care among German clinicians in surgery. *International Journal for Quality in Health Care, 22*, 525-530**.** doi: 10.1093/intqhc/mzq056.

Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd ed.). New York: Guilford Press.

Kokkinos, C. M. (2006). Factor structure and psychometric properties of the Maslach Burnout Inventory-Educators Survey among elementary and secondary teachers in Cyprus. *Stress and Health, 22*, 25-33. doi: 10.1002/smi.1079

Kristensen, T. S., Borritz, M., Villadsen, E., & Christensen, K. B. (2005). The Copenhagen Burnout Inventory: a new tool for the assessment of

burnout. *Work & Stress, 19,* 192-207. doi:
10.1080/02678370500297720

Kyriacou, C. (1987). Teacher stress and burnout: An international review.
Educational Research, 29, 146–152. doi: 10.1080/0013188870290207

Lin, L. P. & Lin, J. D. (2013). Job burnout amongst the institutional
caregivers working with individuals with intellectual and
developmental disabilities: Utilization of the Chinese version of the
Copenhagen Burnout Inventory survey. *Research in Autism Spectrum
Disorders*, *7*, 777-784. doi: 10.1016/j.rasd.2013.03.004

Maslach, C. & Jackson, S. E. (1986). *Maslach Burnout Inventory Manual*
(2nd edition). Palo Alto: Consulting Psychologists Press.

Maslach, C. & Schaufeli, W. B. (1993). Historical and conceptual
development of burnout. In W. B. Schaufeli, C. Maslach & T. Marek
(Eds.). *Professional Burnout*: *Recent Developments in Theory and
Research*. Washington: Taylor and Francis.

Maslach, C., Jackson, S.E., & Leiter, M.P. (1996). *Maslach Burnout
Inventory Manual* (3rd edn.). Palo Alto, CA: Consulting Psychologists
Press.

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual
Review of Psychology*, *52*, 397-422. doi:
10.1146/annurev.psych.52.1.397

Mearns, J., & Cain, J. E. (2003). Relationships between teachers'
occupational stress and their burnout and distress: Roles of coping and
negative mood regulation expectancies. *Anxiety, Stress, and Coping*,
16, 71-82. doi: 10.1080/1061580021000057040

Milfont, T. L., Denny, S., Ameratunga, S., Robinson, E., Merry, S. (2008).
Burnout and wellbeing: testing the Copenhagen Burnout Inventory in
New Zealand teachers. *Social Indicators Research, 89*, 169-177. doi:
10.1007/s11205-007-9229-9.

Odagiri, Y., Shimomitsu, T., Ohya, Y., & Kristensen, T. S. (2004).
Overcommitment and high effort are strongly associated with burnout
among Japanese nurses. *International Journal of Behavioral
Medicine*, *11*(Suppl.), 214.

Papastylianou, A., Kaila, M., & Polychronopoulos, M. (2009). Teachers'
Burnout, Depression, Role Ambiguity and Conflict. *Social*

Psychology of Education, *12*, 295-314. doi:10.1007/s11218-008-9086-7

Pines, A. (1994). The Palestinian Intifiada and Israelis' burnout. *Journal of Cross-cultural Psychology*, *25*, 438-451. doi: 10.1177/0022022194254001

Pines, A. M. (1996). *Couple Burnout*. New York: Routledge

Pines, A. M. (2002). Teacher burnout: A psychodynamic existential perspective. *Teachers and Teaching*: *Theory and Practice*, *8*, 121-140. doi: 10.1037/0033-3204.39.1.103

Pines, A. M. (2004). Why Are Israelis Less Burned Out? *European Psychologist*, *9*(2), 69-77. doi: 10.1027/1016-9040.9.2.69

Pines, A. M., Aronson, E., & Kafry, D. (1981). *Burnout: From Tedium to Personal Growth.* New York: Free Press.

Pines, A. M., & Aronson, E. (1988). *Career Burnout: Causes and cures*. New York: The Free Press.

Platsidou, M., & Agaliotis, I. (2008). Burnout, job satisfaction and instructional assignment related sources of stress in Greek special education teachers. *International Journal of Disability, Development and Education, 55,* 61-76. doi: 10.1080/10349120701654613

Ray, E. B., & Miller, K. I. (1991). The influence of communication structure and social support on job stress and burnout. *Management Communication Quarterly*, *4*, 506-527. doi: 10.1177/0893318991004004005

Schaufeli, W. B & Dierendock., V. D., (1993). The construct validity of two burnout measures. *Journal of Organizational Behaviour, 14*, 631-637. doi: 10.1002/job.4030140703

Schaufeli, W. B., & Enzmann, D. (1998). *The Burnout Campanion to Study and Practice: A Critical Analysis*. London: Taylor and Francis.

Schaufeli, W. B., Leiter, M. P., & Maslach, C. (2009). Burnout: 35 years of research and practice. *Career Development International*, *14*, 204 – 220. doi:10.1108/13620430910966406

Schutte, N., Toppinen, S., Kalimo, R. & Schaufeli, W. B. (2000). The factorial validity of the Maslach Burnout Inventory - General Survey (MBI-GS) across occupational groups and nations. *Journal of*

Occupational and Organizational Psychology, *73*, 53-66. doi: 10.1348/096317900166877

Shirom, A, & Ezrachi, Y. (2003). On the discriminant validity of burnout, depression and anxiety: A re-examination of burnout measure. *Anxiety, Stress & Coping: An International Journal, 16*, 83-97. doi: 10.1080/1061580021000057059

Shirom, A. (2005). Reflections on the study of burnout. *Work & Stress*, *19*, 263–270. doi: 10.1080/02678370500376649

Taris, T. W., Schaufeli, W. B, De Boer, E. M., Schreurs, P. J. G. & Caljé, D. G. (2000). On billijkheid in de arbeidssituatie, gezondheidenziekteverzuim. Hoe ziek is de calculerendeleraar? [Inequity at work, health and absenteeism. How ill is the calculative teacher?] *Gedrag & Gezondheid*, *28*, 123-137.

Taris, T. W., Le Blanc, P., Schaufeli, W. B. & Schreurs, P. J. G. (2005). Are there relationships between the dimensions of the Maslach Burnout Inventory? A review and two longitudinal tests. *Work & Stress, 19*, 256-262. doi: 10.1080/02678370500270453

Weisberg, J., Sagie, A. (1999). Teachers' physical, mental, emotional burnout: impact on intention to quit. *The Journal of Psychology*, *133*, 333-339. doi: 10.1080/00223989909599746

Winwood, P., & Winefield, A. H. (2004). Comparing two measures of burnout among dentists in Australia. *International Journal of Stress Management*, *11*, 282- 289. doi: 10.1037/1072-5245.11.3.282

Maria Platsidou is an Associate Professor of Lifespan Developmental Psychology at the University of Macedonia, Greece

Athena Daniilidou is a PhD Student at the University of Macedonia, Greece

Contact Address: Maria Platsidou: platsidou@uom.edu.gr
Athena Daniilidou: adaniilidou@uom.edu.gr

Enhancing Historical Reasoning: A Strategy Including Formative Assessment with Systematic Continuous Feedback

Sergio Méndez, Felipe Tirado[1]

1) National Autonomous University of Mexico

PLEASE SCROLL DOWN FOR ARTICLE

Enhancing Historical Reasoning: A Strategy Including Formative Assessment with Systematic Continuous Feedback

Sergio Méndez
*National Autonomous
University of Mexico*

Felipe Tirado
*National Autonomous
University of Mexico*

Abstract

Learning History promotes students' reasoning. According to Van Drie & Van Boxtel (2008), historical reasoning involves six elements: substantive concepts, metaconcepts, asking historical questions, using sources, contextualization, and argumentation. Although there are didactic strategies that promote historical reasoning, these do not include systematic continuous feedback using rubrics, which can be useful both in assessing and promoting students' progress and progression of ideas on metaconcepts. This study described the development of the six historical reasoning elements in a strategy that included formative assessment for K8 students. A case study was carried out in Mexico City: four teams of three students were formed according to their knowledge of history, with a single History teacher providing continuous systematic feedback on metaconcepts by using graded rubrics. Results showed that the six historical reasoning elements were developed in different ways and suggested possible methods for use in future didactics.

Keywords: historical reasoning, formative assessment, progression of ideas, rubrics

Promoviendo el Razonamiento Histórico: una Estrategia de Evaluación Formativa con Feedback Sistemático

Sergio Méndez
*Universidad Nacional Autónoma
de México*

Felipe Tirado
*Universidad Nacional Autónoma
de México*

Resumen

La Historia promueve el razonamiento en los estudiantes. El razonamiento histórico involucra seis elementos: conceptos sustantivos, metaconceptos, realizar preguntas históricas, uso de fuentes, contextualización y argumentación (Van Drie y Van Boxtel, 2008). Las estrategias didácticas que promueven dicho razonamiento no consideran la retroalimentación sistemática continua mediante rúbricas, que evalúan progreso y progresión de ideas en metaconceptos. El propósito de este estudio fue describir el desarrollo de los seis elementos del razonamiento histórico en una estrategia que involucró evaluación formativa en estudiantes de segundo grado de secundaria. Se trabajó un estudio de caso en la Ciudad de México: cuatro equipos de tres estudiantes; el profesor de Historia brindó retroalimentación sistemática continua mediante rúbricas calificadas. Los resultados mostraron que los elementos del razonamiento histórico fueron desarrollados de diferente manera y se sugirieron posibles métodos para futuras didácticas.

Palabras clave: razonamiento histórico, evaluación formativa, progresión de ideas, rúbricas.

Hipatia Press
www.hipatiapress.com

Histoy is a subject that has many purposes at school, such as facilitating the comprehension of present times, developing intellectual skills, stimulating extracurricular activities, and acquiring social, aesthetic, and scientific sensibilities (SEP, 2011, p.33). Most importantly, in the process of doing History students develop their reasoning (Lévesque, 2008; Van Drie & Van Boxtel, 2008; Wineburg, 2001). However, despite the potential that History as a subject has, it has lost presence in curricula (Wineburg, 2001) and has been overshadowed in education by Mathematics, language and sciences (Carretero & Castorina, 2010). This is also the case in Mexico, where an evaluation of the subject of History carried out every three years has shown increasingly low scores (SEP, 2010).

In a study on the current state of History teaching in Mexico, Plá & Latapí (2014) stress that, from a psychological point of view, many theoretical and methodological aspects of teaching History are omitted, and that the teaching of this subject from a sociocultural point of view is at an early stage.

Studies on the teaching of History tend to refer to the development of either historical reasoning or historical thinking (Lévesque, 2008; Levstik & Barton, 2011; Wineburg, 2007). Generally speaking, their components are similar and go beyond memorization, a common practice in history teaching that does not demand a high degree of cognitive activity (Carretero & Castorina, 2010). Nevertheless, Van Drie & Van Boxtel (2008) mention that "historical reasoning" emphasizes the students' activities, through which they acquire information of the past and use this knowledge to interpret phenomena of past and present times (p.88); and propose a framework for secondary students which considers the following elements: substantive concepts, metaconcepts, asking historical questions, using sources, contextualization, and argumentation.

Formative assessment is a requirement in the Mexican curriculum for secondary school, which enables teachers to provide feedback to students during the learning process by developing learning strategies (SEP, 2013). Studies on formative assessment show that it can substantially improve students' learning by helping them to understand the learning objectives and

the assessment criteria based on the provided feedback (Black and Wiliam 1998).

Rubrics are used in formative assessment to evaluate students' performance based on learning standards and scales (Mertler, 2001); teachers can rely on rubrics for promoting the learning of content during the educational process (Heritage, 2010). They can also help students judge and comment on their learning, which helps them understand the goal of the rubrics in relation to the established criteria (Sadler, 1989, 1998). Referring to History, there are rubrics in order to assess students' historical reasoning in writing tasks (Monte-Sano & De La Paz, 2012), rubrics for evaluating epistemological instances in historical thinking (Lévesque, 2012) and rubrics that assess historical explanations based on narratives (Levstik & Barton, 2011).

In formative assessment, it is important to take both students' progress and progression into account. The former refers to the acquisition of information that leads students to achieve better grades, while the latter considers both the acquisition of information and the development of the structure of students' ideas (Lee & Shemilt 2003). According to Lee & Shemilt (2003), using metaconcepts or procedural concepts in History teaching is crucial for developing historical thinking, as using only substantive concepts fails to develop the progression of ideas. In addition, Shepard (2009) argues that strategies should include transectional measures in the longitudinal progress, such as assessing various episodes during the educational experience in terms of the progression of ideas. Such progression occurs when students are able to carry their ideas from a concrete level to a critical one (Lévesque, 2012). This can be achieved by providing systematic continuous feedback supported by rubrics in which the progression is evaluated at different levels.

Various studies on teaching History in Mexico have applied formative assessment and show its importance for facilitating the learning process by providing feedback (Plá et al., 2012). Despite the fact that formative assessment refers to the learning process and not only to the end result (Sadler, 1989, 1998; SEP, 2013), the strategies for teaching History which include formative assessment generally disregard the progression of ideas of the metaconcepts, which, according to Lévesque (2008), students need to

appropriate in the process of doing history. In addition, these strategies do not provide continuous systematic feedback, which would make it possible to assess whether students' ideas progressed from a concrete level to a critical one (Lévesque, 2012), and would allow students themselves to assess the progression of their ideas more than once. Finally, there are no strategies in Mexico including formative assessment which take into account the six elements of the historical reasoning framework proposed by Van Drie & Van Boxtel (2008) as a whole.

Purpose of the Study

This study describes a strategy for promoting historical reasoning based on the framework proposed by Van Drie & Van Boxtel (2008) and designed for K8 students in a public secondary school in Mexico, as well as its implementation in a case study in which student's development of the six elements of historical reasoning was observed. The strategy includes formative assessment in order to provide systematic continuous feedback to students based on the rubric criteria which show the progression of ideas of the metaconcepts considered in this study. The ultimate objective of this study is to help students learn to reason and to comprehend history, as well as to provide teachers with a method for designing strategies that promote historical reasoning, without the sole use of memorization.

Strategy Design

This section analyzes the elements of Van Drie & Van Boxtel's (2008) framework, and describes their inclusion in the strategy design.

Substantive concepts

Substantive concepts refer to historical information that can be found in history books, textbooks, films, accounts, and in students' understanding of certain issues, events, phenomena, characters (Lévesque, 2008; Wineburg, 2001), and historical periods (Van Drie & Van Boxtel, 2008). The five historical periods considered in this strategy are those proposed by the Mexican K8 History program (SEP, 2011a). Based on these periods, an open-ended questionnaire was designed to assess the students' degree of historical knowledge by asking them what main historical events happened (see Figure 1).

```
Name:_____  Average:_____

                    HISTORICAL EVENTS QUESTIONS

The following questions intend to find out what information you have about certain historical events.
Please answer the questions below, do not leave any unanswered. These answers will not affect your
school grade. If you have any questions, raise your hand and the teacher will clarify your doubt. Thanks
for your participation.

   1.  - What historical event was the most important in the period from 1960 to 2013?
   2.  - What historical event was the most important in the period from 1920 to 1960?
   3.  - What historical event was the most important in the period from 1850 to 1920?
   4.  - What historical event was the most important in the period from 1750 to 1850?
   5.  - What historical event was the most important in the period from 1550 to 1750?
```

Figure 1. Open-ended questionnaire to assess students' historical knowledge.

Procedural Concepts or Metaconcepts

Procedural concepts or metaconcepts give meaning to the substance of the past by promoting historical inquiry (Lévesque, 2008) and by developing the description and understanding of historical processes (Limón, 2002). In this strategy, we included the metaconcepts mentioned in the History program—causality, progress and decline, primary and secondary sources (SEP, 2011, p. 75)—as well as those representing the past-present-future relation (Pagès, 2003), crucial for developing historical consciousness: Historical significance (importance in the past), Effects in the present, and Envisioning future events.

For this strategy, we designed six rubrics, one for each of the metaconcepts mentioned above (see Table 1), and used them to provide systematic continuous feedback on the progression of the students' ideas in order to encourage students to use them as a learning support by judging their own performance with a critical attitude (Andrade & Du, 2005). Six experts reviewed them and obtained an inter-agreement of 94%, confirming that each rubric was well constructed and that its criteria showed the progression of ideas for each metaconcept considered. In what follows, we describe how the progression of ideas of each metaconcept was assessed. Figure 2 shows the rubric levels describing the progression of ideas for each procedural concept.

Procedural concepts	Rubrics' levels					
	0	1	2	3	4	5
Historical significance	The group fails to mention any event embedded in the period	The group is able to mention only events that affect individuals or their communities, villages, schools, workplaces, towns, etc.	The group mentions, in greater numbers, events which affects a country	The group mentions, in greater numbers, events that have affected several countries (international)	The group mentions, in greater numbers, events have global impact, but does not mention more than five	The group mentions events that have global impact (more than five)
Consequences	The group is not able to mention any kind of consequence related with the historical event	The group is able to mention many benefic consequences or many harmful ones related to the historical event	The group mentions one benefic consequence and one harmful consequence related to the historical event	The group mentions many consequences, and at least, one benefic consequence and one harmful consequence	The group is able to mention many consequences (at least three benefic consequences and three harmful ones) related to the historical event	The group mentions many consequences related to the h historical event (more than three benefic and more than three harmful)
Effects in the present	The group doesn't relate thee effects of t historical event to nowadays	The group is able to relate the effects to the historical event to nowadays considering communities, villages, schools, workplaces, towns, etc.	The group relates the effects of the historical event to the current situation of a country	The group is able to relate the effects of the historical event to the current situation of many countries (internationally)	The group relates the effects of the historical event to the worldwide situation (at least three)	The group is able to relate the effects of the historical event considering completely the worldwide situation (more than three)
Causality	The group doesn't mention any historical character or condition that explains a cause related to the historical event	The group only mentions specific historical characters that explain the causes of the historical event	The group mentions more specific historical characters than the conditions that explain the causes of the historical event	The group mentions more conditions than historical characters that explain the causes of the historical event	The group mentions just the conditions that explain the causes of the historical event	The group mentions just the conditions that explain the causes of the historical event (more than three)
Historical sources	The group doesn't mentions any kind of historical source related to the historical event	The group mentions just secondary sources related to the historical event	The group mentions in secondary sources in great number more than primary ones related to the historical event	The group is able to mention primary sources more than the secondary ones related the historical event	The group just mentions primary sources of the historical event (less than three)	The group is able to mention just primary sources of the historical event (more than three)
Envisioning future events	The group doesn't envision any future event based on the effects of the historical event	The group is able to envision future events based on the effects of the historical event considering communities, villages, schools, workplaces, towns, etc.	The group envisions more future events, based on the effects of the historical event, considering a country	The group envisions more future events, based on the effects of the historical event, considering many countries (internationally)	The group envisions more future events, based on the effects of the historical event, considering worldwide situation (no more than three)	The group is able to envision just future events, based on the effects of the historical event, considering the worldwide situation (more than three)

Figure 2: Levels of the rubrics that show the progression of ideas for each procedural concept

Historical significance refers to the individual's capacity of identifying the most significant events (Lomas, 1990). The progression of ideas in this procedural concept was assessed by asking students to differentiate between events that involve a single person or place (Level 1 in the rubric), and those involving a larger number of people worldwide (Level 5 in the rubric); this assessment was based on the criteria of quantity (Partington, 1980).

Consequences attend the understanding of historical consequences in the past. The rubric assessed the students' progression of ideas regarding their understanding of the fact that historical events have many beneficial and many harmful consequences (Level 5), and do not just have one beneficial or one harmful consequence (Level 1). When students integrate both perspectives, a thoughtful and critical history is constructed (McCarthy, 1998); the sense of history is restored; and a secular history, which involves progress and decline, is articulated (Le Goff, 2005).

Effects in the present refer to understanding the effects of historical events on the present. Carretero & Montanero (2008) state that by understanding the present times, collective memory is extrapolated and mental representations are articulated. The Effects in the present rubric assesses the students' progression of ideas regarding this understanding from the individual (Level 1) to a global scale (Level 5) (Chesnaux, 2009).

Causality refers to understanding why events happened and what circumstances contributed to their origin (Montanero & Lucero, 2011). The progression of ideas, assessed with the Causality rubric, involves understanding that historical events were not originated only by historical characters (Level 1), that is a common sense explanation (Halldén, 1998); but also by considering the historical context (Level 5), because students tend to have very limited or mistaken conceptions of this one (Wineburg, 2001).

Historical sources involve understanding documents, objects, images, etc. that provide relevant information of historical events (Prats, 2001). While the use of Historical sources is one of the elements in Van Drie and Van Boxtel's (2008) historical reasoning framework and will be discussed in the following section, here, Historical sources are considered as a metaconcept in order to observe the students' progression of ideas. The

Historical sources rubric assesses students' discernment between sources that were produced when the events emerged (primary sources) (Level 5), and the reflections or comments that have been made based on them (secondary sources) (Level 1) (Prats, 2001). Primary sources have been privileged in the analysis of history (Lévesque, 2008) and are the main sources used in the classroom to understand history (Prieto, Gomez & Miralles, 2013).

Envisioning future events drives students to imagine forthcoming events, based on the effects of a past event in present times. The progression of ideas, assessed by the Envisioning future events rubric, goes from imagining future events that consider one person (Level 1) to those that consider many people worldwide (Level 5). To envision is not to determine, but to represent and imagine (Staley, 2002), and it needs to be based on evidence by joining past and future times (Staley, 2007).

While all these metaconcepts are closely related, we follow Van Drie and Van Boxtel (2008) historical reasoning framework in considering them separately, by using rubric levels to obtain an objective assessment of the progression of ideas on each metaconcept.

Asking Historical Questions

Asking historical questions shapes and promotes inquiry by working with procedural concepts (Counsell, 2000). According to Levstik & Barton (2011), this inquiry should be a disciplined one that teaches students what to ask and how to answer historical questions by finding information, evaluating sources, and integrating conflicting explanations to provide an interpretation.

Lévesque (2008a) has underlined the importance of students searching for information in digital environments, especially in History. If students search only for information supported by the textbook, they construct its contents as absolute truths (Carretero, Jacott & López-Manjón, 2002).

To implement asking historical questions in our strategy, we designed six historical questions based on the metaconcepts described above (see Figure 3).

Figure 3. KronoTop template.

The teacher asked the students to answer six questions per historical period, one by one, by looking for responses in the history textbook and in the web. Because asking historical questions refers to promoting inquiry in students, we first decided to teach them what types of questions to ask in order to see whether they asked other questions during the learning process by responding to the ones constructed by the teacher with the support of the rubrics. The questions were embedded into five PowerPoint templates designed for the purpose of this strategy, each of which represents one of the five historical periods considered in this strategy. Use of sources and Contextualization are also integrated into these templates (see Figure 3).

Use of Sources

Nowadays, technology has enabled us to search and find primary sources quickly and simply (Lee, 2002). Objects, images, and all kinds of documents can be found in order to obtain a wide range of information (Van Drie & Van Boxtel, 2008). Studies of History teaching examine the use of digitized primary sources and mention positive perspectives towards their use (Hicks, Doolittle & Lee; 2004; Waring & Torrez, 2010).

To implement use of sources in our strategy, students searched for primary digitized sources in the web in order to illustrate the historical events they had selected by making a collage and placing it in the appropriate space in a Power Point template designed specifically for this strategy (see Figure 3). Before this task, students needed to identify the difference between primary and secondary sources, as shown by the progression of ideas of Historical sources as a metaconcept.

Contextualization

Contextualization is defined as the competence to place a historical phenomenon, an object, an argument, a text, or a drawing into a social, spatial, and temporal context in order to describe, explain, compare, and evaluate it (Van Drie & Van Boxtel, 2008). According to a study by Shemilt (1983), adolescents have difficulties in making sense of history by trying to place historical events. By working with Contextualization, the intention is for students to interpret and understand historical events, based on their own historical time (Wineburg, 2007). In order to promote Contextualization, we indicated the historical period being researched in a space at the top of the PowerPoint template, thus representing time (Kronos). To represent place/space (Topos), we included a space below it, in which students were asked to locate the historical events of the period in a map (see Figure 2). For these reasons, we called the template KronoTop.

Argumentation

The Use of sources is related to argumentation because arguments are based on documented evidence (Perfetti et al., 1994) that supports the students' claims (Barton & Levstik, 2004). Van Drie & Van Boxtel (2008) also stress the importance of argumentation, considering it one of the six elements of historical reasoning. However, while they refer to the quality of argumentation, they do not mention its assessment. Therefore, Toulmin's model of argumentation is useful because it is embedded into the general domain models of argumentation where the quality of arguments can be assessed inside or outside the scientific field (Sampson & Clark, 2008). In addition, this model is useful in historical reasoning because it stresses that evidences play an important role in an argument structure (Toulmin, 2003), and because it has been found to be useful in applying historical reasoning to ill-defined problems (Voss, 2006). Based on Toulmin's model, Simon, Erduran & Osborne (2006) made a distinction between argument and argumentation. They define argument as the set of statements, data, guarantees, and backings that are involved in the conformation of the pronounced argument, while argumentation is defined as the process of joining these components. The use of rebuttals is a complex skill that allows students to argue which argument is better by demonstrating a greater commitment and ability to integrate original and alternative claims (Kuhn, 1991).

Erduran, Simon & Osborne (2004) generated a method for analyzing the quality of argumentation in small groups of students, taking the above-mentioned components into account. They assessed the quality of argumentation based on the nature and presence of the rebuttals emitted by the students involved in the argumentation. A low level quality of argumentation indicates an opposition among students consisting of unrelated counterarguments that do not challenge the validity of the evidence or of the justifications that were offered previously in an understanding of refutation. However, when there is a rebuttal that defies the pieces of evidence (statements, guarantees or backings), the argumentation is considered high quality (see Table 1).

Table 1:
Analytical Framework used for assessing the Quality of Argumentation proposed by Erduran, Simon & Osborne (2004, p. 928).

Level 1	Level 1 argumentation consists of arguments that are a simple claim versus a counter-claim or a claim versus a claim.
Level 2	Level 2 argumentation has arguments consisting of a claim versus a claim with either data, warrants, or backings but do not contain any rebuttals.
Level 3	Level 3 argumentation has arguments with a series of claims or counter-claims with either data, warrants, or backings with the occasional weak rebuttal.
Level 4	Level 4 argumentation shows arguments with a claim with a clearly identifiable rebuttal. Such an argument may have several claims and counter-claims.
Level 5	Level 5 argumentation displays an extended argument with more than one rebuttal.

From a sociocultural perspective, argumentation is essential for learning science and its appropriation is promoted by working within communities of practice (Kelly & Chen, 1999). Students develop argumentation by discussing topics while they are embedded in a dialogic process (Mortimer & Scott, 2003) in which they manage to externalize their thinking, transcending the intra-psychological act and staying in the inter-psychological one (Vygotsky, 1978); and in which the teacher promotes collaboration and participants can provide social support or scaffolding, which generates the principle of Proximal Development Zone (Vygotsky, 1986).

Coffin & O'Halloran (2009) conclude that the argumentation subject has changed from being a "combat adversary" to a "dialogic exchange" (p. 302). For this reason, our strategy was designed to have the students reach a consensus, instead of pointing at the student who made the best argument.

Case Study

The strategy designed to enhance the six elements of historical reasoning as a whole was implemented in a case study with a group of K8 students in a public secondary school in Mexico. The strategy included both providing formative assessment with systematic continuous feedback on metaconcepts using rubrics and the assessment of the quality of argumentation in the students' dialog.

The strategy designed to enhance the six elements of historical reasoning as a whole was implemented in a case study with a group of K8 students in a public secondary school in Mexico. The strategy included both providing formative assessment with systematic continuous feedback on metaconcepts using rubrics and the assessment of the quality of argumentation in the students' dialog.

This strategy was designed and implemented with secondary students because the curriculum in Mexico stresses the importance of developing historical thinking in primary and secondary schools (SEP, 2011a), and because the framework on which we based the strategy design is directed at secondary students (Van Drie and Van Boxtel, 2008). K8 grade was selected because in K7 grade, the first secondary grade in Mexico, History is not taught in Mexican public schools (SEP, 2011a).

The case study method was selected because it allows the strategy to be considered in a real environment (Yin, 1994). The public secondary school selected for the case study has a very low degree of marginalization and high scores in most subjects, according to SEP (2010).

In order to select the participants for the strategy, an open-ended questionnaire referring to substantive concepts (Figure 1) was designed based on the five historical periods mentioned by the Ministry of Education for K8 grade (SEP, 2011a) and was applied to the K8 students of the chosen secondary school. Two judges graded the questionnaires without any discrepancies.

Based on their grades, four students who obtained more than 80% correct answers were selected (high performance); four who obtained between 60% and 79% correct answers (average performance); and four who obtained less than 59% (low performance). The twelve participants selected—six males

and six females—were thirteen years old and had not interrupted their studies nor taken extra classes beyond the ones received at school. Based on this selection, four teams were created (A, B, C and D) with three students in each one (medium — average — high), in order to consider the principle of Zone of Proximal Development (Vygotsky, 1986), in which the more advanced students help the less advanced. The teacher worked with each team separately after an informed consent form based on the ethics code (APA, 2010) was obtained from the students and their parents.

Once each team met with the teacher, he asked each student to send him the description of five historical events they considered to be of great importance for the first period (1960-2013) via e-mail. Subsequently, each of the four teams gathered separately with the teacher and read the first question embedded in the KronoTop template: "What were the most significant events in this period". The students were asked to reach a consensus and to write their responses below the first question—a task designed in order to consider the metaconcept of Historical significance.

Subsequently, the teacher asked each team to choose one of the significant events they agreed on in order to answer, by consensus, the following five questions in the KronoTop template which referred to the metaconcepts of Consequences, Effects in the present, Causation, Evidences, and Envisioning future events, respectively. For the purpose of these tasks, the teacher allowed the students to search for information on the web and in the History textbook, which allowed them to develop disciplined inquiry, that is, to learn what types of questions to ask and how to answer historical questions (Levstik & Barton, 2011), Our interest lay in providing students with historical questions to see whether they posed other questions based on the previous one and on the systematic continuous feedback provided by the rubrics.

The teacher then asked each team to choose, by consensus, two to five representative images of the period using an image web browser, and to make a collage in the KronoTop template ("Use of sources"). Subsequently, the teacher asked each team to locate the historical events that occurred in the historical period being discussed in a map and to copy it into the template ("Contextualization"). The purpose of asking students to reach

consensus in their answers was to be able to assess the quality of their argumentation.

The teacher then proceeded to grade the six rubrics based on the teams' responses in the KronoTop template and show them to each team before starting work on the tasks for the next period (1920-1960), thus providing continuous systematic feedback (formative assessment). Students were able to observe the grades they achieved during the strategy (progress) and the structure of their ideas (progression); they argued about the quality of their responses based on the rubrics' criteria and if they did not understand these, the teacher explained them in order to motivate the students to discuss what they needed to do to increase their scores for the next historical period.

Each team repeated the same process until they completed the five periods considered in the strategy. After each period, students' responses in KronoTop were used in order to assess the reliability of the rubrics. Two judges obtained a kappa coefficient of 0.86, which is considered a very acceptable inter-agreement according to the values indicated by Abad et al. (2011). All sessions were videotaped and at the end of the strategy, the teacher asked the students to answer the open-ended questionnaire again in order to determine whether students were able to relate events to the periods in which they occurred.

Development of the Elements of Historical Reasoning in the Case Study

In this case study, the elements of historical reasoning proposed by Van Drie & Van Boxtel (2008) were developed in different ways, as will be analyzed in this section. This development was related with the systematic continuous feedback provided to the students, which led to the development of the student's disciplined inquiry and argumentation. While the latter was generated, its quality was low.

The development of the metaconcepts can be observed in the progress and progression of ideas, shown in the rubrics, during the learning process that comprehended the five historical periods mentioned by the Ministry of Education (SEP, 2011a). As mentioned above, rubrics were designed to facilitate progress and progession on metaconcepts. Level 5 in the rubrics

refers to increasing the number of events, showing progress, and Level 4 refers to the change of the structure of ideas, showing progression.

The most significant events of each period selected by the teams are shown in Table 2.

Table 2:
Events selected by team consensus.

HISTORICAL PERIODS	*1960-2013*	*1920 – 1960*	*1850 – 1920*	*1750 – 1850*	*1550 -1750*
Team A	AIDS emergence	Sputnik launch	Telephone invention	Industrial Revolution	Protestant Reformation
Team B	Technological changes	Polio vaccine	Second Industrial Revolution	Origin of Species Publication	Spices exchange
Team C	Apolo XI	Penicillin vaccine	Second Industrial Revolution	Industrial Revolution	Renaissance
Team D	Mexico City's earthquake in 1985	Second World War	First World War	Enlightenment	Newton's physics theory

Table 3 refers to the teams' progress by showing the scores of the first period (P1) in which feedback was not provided and the average of the other four (P4) in which feedback was provided by the teacher. The mean increased in all cases, suggesting that the students' progress regarding the metaconcepts was achieved due to the systematic continuous feedback provided by the teacher using the rubrics.

Table 3:
Scores for the four teams.

	Events		Consequences		Effects		Causes		Sources		Envisioning future events	
	P1	**P4**	**P1**	**P4**	**P1**	**P4**	**P1**	**P4**	**P1**	**P4**	**P1**	**P4**
Team A	4	5	3	4.5	1	4.5	0	4.25	1	4.75	2	4.5
Team B	4	4.75	3	4	4	4.75	1	4.75	1	4.5	2	5
Team C	3	4.75	2	4	1	4.75	3	5	1	4.75	2	5
Team D	3	4.75	2	3.75	1	4.5	2	3.5	1	5	0	4.5

Referring to Historical significance, all teams understood the criteria by mentioning events that involve people worldwide (progression), and three of them reached the Level 5 of the rubric which involves mentioning more than three events based on this criteria (progress). Team B was the only one that did not reach Level 5 for the period from 1550 to 1750, having mentioned just three events that involve people worldwide: spice exchange, slave trade, and the Independence of the Thirteen Colonies in North America.

Similar results were observed regarding Consequences. All teams showed progression in their ideas because all of them comprehended that historical events produce both beneficial and prejudicial consequences, reaching Level 4 in the rubric. Progress was not reached completely: teams C and D had difficulty citing more than three benefits and more than three damages resulting from the Renaissance and Newton's theory of physics (Level 5). For example, in the case of the Renaissance, team C mentioned more than three benefits and only one damage reached in consensus: the Church lost believers.

As seen in Figure 4, working with the metaconcept of Effects in the present, progress and progression of ideas were completely accomplished

(Level 5). They could easily relate more than three past events that affect people worldwide with today, based on what they had experienced in their own lives, without searching for information neither in the web nor in the History textbook.
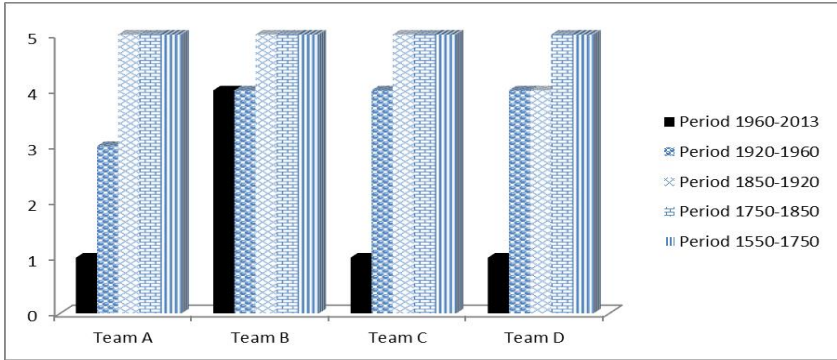

Figure 4 "Effects in present times" progression of ideas in each team for each period.

Referring to Causality, students' scores declined on three occasions, as seen in Figure 5: team A when working on the telephone invention event (third period), and team D when working on WWI and Newton's physics theory (third and fifth periods, respectively). These observable decreases were due to the fact that the teams gave more importance to historical characters than to the historical context to explain Causality (see Levels 2 and 3 in the rubrics); therefore, neither progression nor progress was observed.
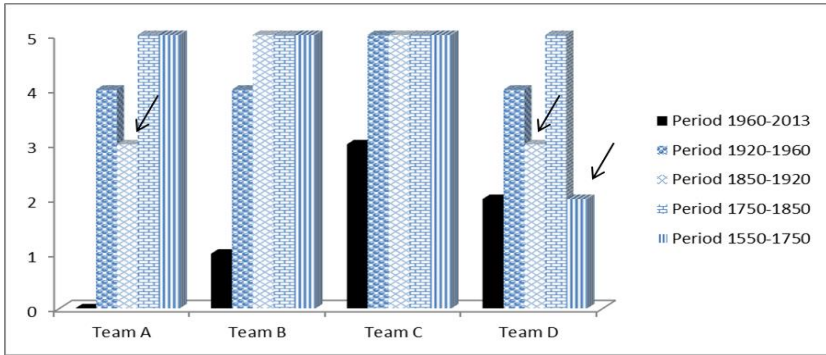
Figure 5. "Causes" progression of ideas in each team for each period.

All teams showed a progression of ideas regarding Historical sources (see Figure 6). Team B was the only one with a score decrease when working with the last period, because it did not mention enough primary sources to reach the maximum rubric score (Level 5—more than three primary sources).
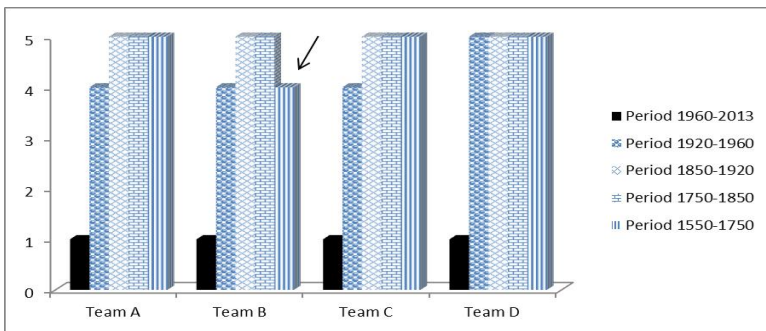

Figure 6. "Historical sources" progression of ideas in each team for each period.

There was progression in all cases when working with the metaconcept of Envisioning future events, evident in the comprehension of imagining events that involve people worldwide based on the Effects in the present. Progress was not completely accomplished because there was a decrease in

Team A's score in the last period (Protestant Reformation) due to not envision more than three events that involved people worldwide (Level 5). All teams referred to future events without searching for information, like they did when working with the Effects in the present metaconcept.

Using rubrics to provide systematic continuous feedback was useful in developing and assessing progress and progression on the six metaconcepts considered in this strategy. It led students to pose questions based on the rubrics' criteria in order to achieve better grades, change the structure of their ideas, and generate argumentation with the members of each team.

The fact that students posed questions is related to the way in which the component of Asking questions (disciplined inquiry) was incorporated: students learnt what kind of historical questions they needed to answer by reading the ones embedded in KronoTop, and they learnt how to answer them by looking for information on the web and in the History textbook. The following dialogue shows a disciplined inquiry made by the students when they tried to answer the question "What were the most significant events in this period (from 1550 to 1750)?":

> S1: Who did this historical event affect? The whole world.
> S2: [reading the History textbook] The Independence of the Thirteen Colonies?
> S3: But, that event is not worldwide. Well, OK.
> S2: From who did the Thirteen Colonies get their independence?
> S3: United States. Only.
> S2: ¿And it didn't affect the whole world? ¿This event just affected them?
> S1: Perhaps it affected [the whole world] with the economy.
> S2: [reading the History textbook] As you wish, but the Independence of the Thirteen Colonies was in 1776.
> S1: In that case no, because the period is from 1550 to 1750.
> S3: Ok, keep on searching

In this dialog, students looked for information in the History textbook in order to find what the most significant events in that period were. They also asked other questions based on the rubrics' criteria as they looked for events that had impacted people worldwide and realized that the historical event of the Independence of the Thirteen Colonies did not fall into the period they

were considering. By answering questions, posing others, and finding information, students developed discipline inquiry as established by Levstik & Barton (2011). Other kinds of inquiries were observed in the students' discussions among themselves. These inquiries were also based on the rubrics' criteria, which referred to the progression of ideas for the metaconcepts included in the strategy.

As well as leading the students to pose questions, the continuous systematic feedback elicited argumentation. As the previous dialog showed, the members of each team were able to build arguments based on the rubrics' criteria for developing progress and progression on the metaconcepts considered in this strategy. In addition, argumentation was elicited by answering the questions in KronoTop and by looking for information on the Web and in the History textbook.

Videotapes of the students' argumentation process for reaching consensus on each procedural concept were transcribed. The analysis of the quality of argumentation was done based on the levels suggested by Erduran, Simon & Osborne (2004) (Table 1). Team A was chosen for this analysis because its score on the metaconcept of Causality improved from 0 to 4 after feedback was provided and it had the highest number of total responses in the strategy.

Team A's results showed that the largest number of arguments were at level 1; very few arguments were generated at levels 2 and 3; and arguments at levels 4 and 5 never appeared (see Figure 7). Two judges obtained a very good level of inter-agreement (k = 0.82) according to the values expressed by Abad et al. (2011).
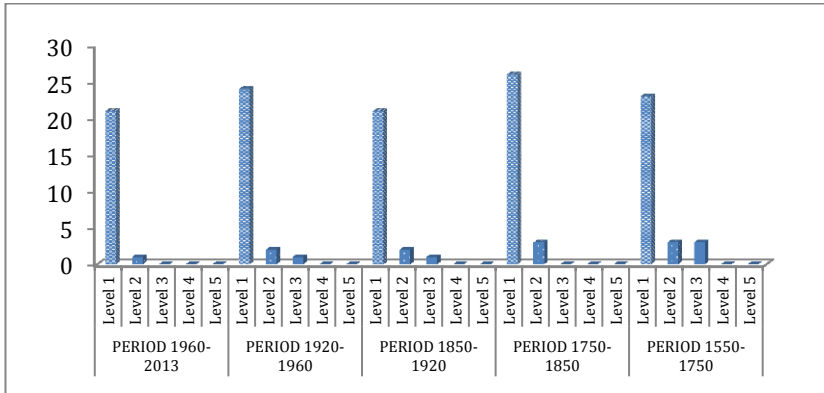
Figure 7. Quality of argumentation in team A by historical period.

The following dialog shows an example in which the consequences of the historical event (technological changes) from 1960 to 2013 were discussed by the students:

> S1: More communication, the communication was a bit easier.
> S2: The communication…is more…
> S1: The communication using signs of…
> S3: But there were more robberies.
> S1: What?
> S3: But there were more robberies.
> S1: Yes, but that we can write as damage.

In the conversation, a weak rebuttal is offered by student 3 (S3) who argues that the consequence was not the one put forward by S1, but the increase in robberies. It is considered a weak rebuttal because it was not taken into account for the consensus in the end, and student 3 (S3) did not back up or strengthen his claim in order to substitute the first claim with his own.

Argumentation and disciplined inquiry played an important role in various tasks of the strategy: answering the questions about the metaconcepts by reaching consensus, looking for digitized sources ("Use of

sources") in the Web to make a collage in KronoTop illustrating the period, and locating the significant events in a map ("Contextualization"). Throughout these tasks, students discussed, reached consensus, looked up information, and posed questions based on the previous answers in KronoTop.

Concerning "Use of Sources" and "Contextualization", figure 8 shows a KronoTop template, in which historical events are illustrated with digitized sources and through the location of the events in the maps; according to the period from 1750 to1850.
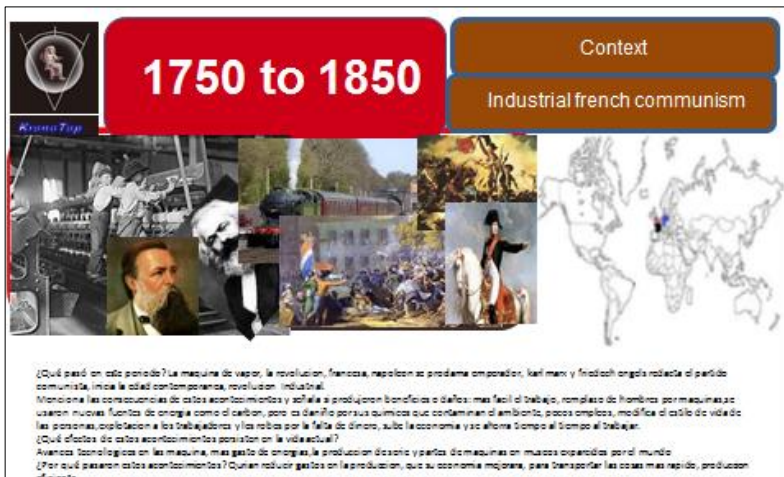

Figure 8. Use of digitized sources and spatial location of historical events in the period from 1750 to 1850.

With regard to substantive concepts, the open-ended questionnaire applied at the end of the strategy showed significant differences from the one which was applied before (T = 2, n = 11, p <0.01), according to the values offered by Triola (2009); just eleven students responded the questionnaire at the end, because one of them dropped out of school. The History teacher graded student's answers in the questionnaires, and found that they corresponded with what the History textbook states. This indicates that

students were able to expand on previously acquired information. The tasks that could help to this result were: sending the historical events to the teacher via e-mail, working with KronoTop templates because the period was explicitly written at the top of them, and answering the first question by consensus, which referred to Historical significance.

Conclusions

A strategy including formative assessment with systematic continuous feedback based on the progression of ideas of metaconcepts. was implemented in a case study with K8 grade students and in general terms, the six elements of historical reasoning proposed by Van Drie & Van Boxtel (2008) were developed in the four teams that participated in it. While the results are not generalizable, they help to understand what tools may aid students in developing their historical reasoning and what adaptations might be implemented in future strategies, as will be discussed in this section.

During the learning process, the results showed that the progression of ideas for each metaconcept was different, which is consistent with the observations made by Lee & Shemilt (2003) who found that the procedural concepts or metaconcepts are not developed in parallel, but differ in their appropriation, which underscores the importance of the context of each historical event. Another explanation for the few decreasing scores in the rubrics may be the diversity and complexity of historical events, which entail a different level of analysis. Finally, the fact that students reached higher scores in the rubrics might suggest that, as Lévesque (2012) proposes, they generated more critical and realistic thinking.

The lowest grades obtained by the teams were in the Causality rubric, reflecting the non-progression of their ideas. Students tended to prioritize historical characters in order to explain the origin of the historical events, which is consistent with the observations of Halldén (1998). It would therefore be necessary, in future strategies, to write in the Causality rubric the type of contexts (economic, social, cultural, etc.) that might explain the origin of a historical event, instead of just mentioning "conditions".

Students should be motivated to search for more information about the farthest historical events because most descriptions in which progress was

not reached by the teams referred to the farthest period (from 1550 to 1750). Also, in future strategies, students should be asked to support their answers with evidence, especially when working with the metaconcepts of Effects in the present and Envisioning future events, as Staley (2007) argues, because teams did not search sources in the web nor in the History textbook by answering the questions of these metaconcepts during the strategy.

Although students failed to search for information in order to support their responses, specifically for the metaconcept of Effects in the present, all teams reached the highest scores on the rubrics for all the periods considered in the strategy. This suggests that students managed to concatenate significant events up to the present and articulate mental representations of the current time, which demonstrates progression in terms of this concept (Carretero & Montanero, 2008) and which constitutes an advance in the field, which, as Muñoz & Pagés (2012) point out, is very necessary.

Considering Asking historical questions, the dialog showed that the KronoTop templates and the rubrics were useful to the students because they were able to learn what to ask by responding the questions in the KronoTop templates and how to answer historical questions by searching the information in the web and in the History textbook, as discipline inquiry mention (Levstik & Barton, 2011). Likewise, students posed other questions based on the rubrics' criteria. Because students used technology in order to find information, the content of textbooks was no longer regarded as absolute truth, an effect observed by Carretero, Jacott & López-Manjón (2002).

Digital sources were used as illustrations when they were considered significant in supporting the occurrence of the events. Once students have the ability to distinguish between primary and secondary sources, they are able to use them as evidence for the creation of hypothesis and interpretations, thus achieving better contextual thinking (Dickinson & Lee, 1980; Wineburg, 2007).

Regarding the element of Contextualization, students related the events temporally and geographically using the KronoTop template. With regard to argumentation, the framework proposed by Erduran, Simon & Osborne (2004), based on Toulmin's model for the oral evaluation of the quality of argumentation in students, was useful for the subject of History and

confirmed with small groups. Results showed that the quality of argumentation in the assessed team was very low, consistent with the results of Van Drie et al. (2006), who observed that students cited several arguments to support their claim without mentioning rebuttals. This suggests that although in this strategy students were asked to reach consensus setting aside debate, as Coffin & O'Halloran (2009) suggest, it might be necessary to design tasks that involve both debate and consensus in order to see whether students consider contradictory statements to generate rebuttals, and at the end have the opportunity to deliberate and reach consensus.

Despite having created teams in communities of practice with different levels of expertise, as Lave & Wegner (1991) suggest, students only managed to co-construct arguments, and a high quality of argumentation was not achieved. Therefore, it will be necessary to develop argumentation by implementing any of the following strategies: direct explanation through instruction, structured tasks and modeling as Mason (1996) suggests; the teacher can ask questions in order to make students provide arguments in their answers, as Simon, Erduran & Osborne (2006) propose; or constructing and validating rubrics that show levels in which the progression of argumentation could be noted in order to change students' ideas over time, as Smith, Wiser, Anderson, & Krajcik (2006) suggest. This last point is of particular importance because there is little research regarding the progression of argumentation (Duschl, Schweingruber, & Shouse 2007). In general, the results showed that the strategy was useful in helping students reach higher scores on the substantive concepts questionnaire at the end of the strategy; helping students look for the information to answer historical questions and ask historical questions in order to reach higher scores on the metaconcepts rubrics (to carry out disciplined inquiry); helping students be able to find digital sources by discerning between primary and secondary sources; helping students contextualize historical events by matching their location in a map with the digital sources; and eliciting argumentation among the students, who discussed the criteria of each metaconcept rubric in order to reach a consensus on each question. The KronoTop templates were useful to find digitized sources and to locate historical events, as well as to develop their historical reasoning by developing its six elements. In future

research, it will be important to evaluate the strategy in an entire classroom and analyze the results in order to make further generalizations.

References

Abad, F., Olea J., Ponsoda, V., y García, C. (2011), *Medición en ciencias sociales y de la salud*, España: Síntesis.

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10 (3), 1-11.

APA (2010). *Ethical Principles of Pyschologists and Code of Conduct* http://www.apa.org/ethics/code/principles.pdf (Accessed 2013-03-14).

Barton, K. C., & Levstik, L. S. (2004). *Teaching history for the common good*. Mahwah, NJ: Erlbaum

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*, 5(1), 7-74.

Carretero, M., & Castorina, J. (2010). *La construcción del conocimiento histórico. Enseñanza, narración e identidades*. Argentina: Paidós.

Carretero, M., Jacott, L., & López-Manjón, A. (2002). Learning history through textbooks: are Mexican and Spanish students taught the same story?. *Learning and Instruction*, 12 (6), 651-665. doi:10.1016/S0959-4752(01)00036-6

Carretero, M., & Montanero, M. (2008). Enseñanza y aprendizaje de la Historia: aspectos cognitivos y culturales. *Cultura y Educación*, 20 (2), 133-142.

Chesnaux, J. (2009). *¿Hacemos tabla rasa del pasado?* México, DF: Siglo XXI.

Coffin, C., & O'Halloran, K. A. (2009). Argument reconceived?. *Educational Review*, *61*(3), 301-313. doi:10.1080/00131910903045948

Counsell, C. (2000) 'Historical knowledge and historical skills: a distracting dichotomy', in J. Arthur and R. Phillips (eds) *Issues in History Teaching*. London: Routledge: 54–71.

Dickinson, A.K., & Lee, P.J. (1980). *History Teaching and Historical Understanding*. Heinemann: London, 1-19.

Duschl, R. A., Schweingruber, H. A., & Shouse, A. E. (Eds.). (2007). *Taking science toschool: Learning and teaching science in grades K-8*. Washington, DC: National Academies Press.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science education*, *88*(6), 915-933. doi:10.1002/sce.20012

Halldén, O. (1998). Personalization in historical descriptions and explanations. *Learning and Instruction*, 8 (2), 131-139. doi:10.1016/S0959-4752(97)00012-1

Heritage, M. (2010). Formative Assessment and next-generation Assessment Systems: Are We Losing an Opportunity?. *Council of Chief State School Officers*, 1-20.

Hicks, D., Doolittle, P., & Lee, J. (2004). Social Studies' Teachers Use of Classroom Based and Web-Based Historical Primary Sources. *Theory and Research in Social Education*. 32, (2), 213-247. doi:10.1080/00933104.2004.10473253

Kelly, G. J., & Chen, C. (1999). The sound of music: Constructing science as sociocultural practices through oral and written discourse. *Journal of Research in Science Teaching*, *36*(8), 883-915. doi:10.1002/(SICI)1098-2736(199910)36:8<883::AID-TEA1>3.0.CO;2-I

Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.

Lave, J., & Wenger, E. (1991). *Situated learning, legitimate peripheral participation*. New York: Cambridge University Press.

Lee, J.K. (2002). Digital history in the history/social studies classroom. *History Teacher*, 503-517. doi:10.2307/1512472

Lee, P. & Shemilt, D. (2003). A scaffold, not a cage: progression and progression models in history. *Teaching History*, 113, 13-22.

Le Goff, J. (2005). Pensar la historia. Modernidad, presente, progreso. Mexico: Paidós.

Lévesque, S. (2008), *Thinking Historically*. *Educating Students for the Twenty-First Century.* Canada: University of Toronto.

Lévesque, S. (2008a). The impact of digital technologies and the need for technological pedagogical content knowledge. In T. Di Petta (Ed), *The*

Emperor's New Computer: ICT, Teachers and Teaching, (pp. 17-28). Sense Publishers: Rotterdam/Taipei

Lévesque, S. (2012). Thinking the War of 1812: Notes on Student's Historical Ideas. *Canadian Issues*, 41-47.

Levstik, L., & Barton, K. (2011). *Doing history. Investigating with children in elementary and middles schools.* NY:Routledge.

Limón, M. (2002). Conceptual change in history. In M. Limón and L. Mason, L. (Eds) *Reconsidering conceptual change: Issues in theory and practice* (pp. 259-289). Springer: Netherlands.

Lomas, T. (1990). *Teaching and Assessing Historical Understanding*, London: The Historical Association.

Mason, L. (1996). An analysis of children's construction of new knowledge through their use of reasoning and arguing in classroom discussions. *Qualitative Studies in Education*, 9(4), 411–433.

McCarthy, T. (1998), *La teoría crítica de Jürgen Habermas.* Madrid: Tecnos.

Mertler, C. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research &Evaluation*, 7 (25), 1-9.

Montanero, M., & Lucero, M. (2011). Causal discourse and the teaching of history. How do teachers explain historical causality?. *Instructional Science*, 39(2), 109-136. doi:10.1007/s11251-009-9112-y

Monte-Sano, C., & De La Paz, S. (2012). Using writing tasks to elicit adolescents' historical reasoning. *Journal of Literacy Research*, *44*(3), 273-299. doi: 10.1177/1086296X12450445

Mortimer, E., & Scott, P. (2003). *Meaning Making In Secondary Science Classroomsaa*. McGraw-Hill International.

Muñoz, E., y Pagès, J. (2012). La relación pasado-presente en la enseñanza de la historia para la enseñanza de la historia en la educación secundaria obligatoria catalana. *Clío & Asociados*, 11-38.

Pagès, J. (2003). Ciudadanía y enseñanza de la historia. *Reseñas de Enseñanza de la Historia*, 1, 11-42.

Partington, G. (1980). *The Idea of an Historical Education*. Windsor, UK: NFER Publishing.

Perfetti, C. A., Britt, M. A., Rouet, J. F., Georgi, M. C., & Mason, R. A. (1994). How students use texts to learn and reason about historical

uncertainty. *Cognitive and instructional processes in history and the social sciences*, 257-283.

Plá, S. y Latapí, (2014). La construcción de un campo de investigación: la enseñanza de la historia en México. En S. Plá y J. Pagés (coord.), *La investigación en la enseñanza de la historia en América latina* (pp.193-216). México: UPN.

Plá, S., Rodríguez, X. y Gómez, V. (2012) (coord.). Miradas diversas a la enseñanza de la historia. México: UPN.

Prats, J. (2001). Enseñar Historia: *Notas para una didáctica renovadora. Mérida: (ES).*

Prieto, J., Gómez, J., y Miralles, P. (2013). El uso de fuentes primarias en el aula y el desarrollo del pensamiento histórico y social. Una experiencia en Bachillerato, *Clío*, (39), 34-14.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science*, *18*(2), 119-144.

Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in education*, *5*(1), 77-84.

Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, *92*(3), 447-472. doi:10.1002/sce.20276

SEP. (2010). *Boletín informativo de ENLACE*. Retrieved from http://www.enlace.sep.gob.mx/gr/

SEP (2011). Enseñanza y aprendizaje de la Historia en Educación Básica. Retrieved from http://basica.sep.gob.mx/HISTORIA%20web.pdf.

SEP. (2011a).*Competencias Digitales. Programas de Estudio 2011. Guía para el maestro.* Retrieved from http://basica.sep.gob.mx/dgdc/sitio/pdf/inicio/matlinea/2011/Historia_SEC.pdf

SEP. (2013). *El enfoque formativo de la evaluación. SERIE. Herramientas para la Evaluación en Educación Básica*. Retrieved from http://basica.sep.gob.mx/dgdc/sitio/pdf/serieherramientas/C1*WEB* 2013.pdf

Shemilt, D. (1983). The devil's locomotive. *History and Theory*, 22 (4), 1-18.

Shepard, L. (2009). Commentary: Evaluating the Validity of Formative and Interim Assessment. *Education of Measurement*, 28 (3), 32-37. doi:10.1111/j.1745-3992.2009.00152.x

Simon, S., Erduran, S., & Osborne, E. (2006). Learning to Teach Argumentation: Research and development in the science classroom. *International Journal of Science Education*, 28 (2-3), 235-260.

Smith, C. L., Wiser, M., Anderson, C. W., & Krajcik, J. (2006). Implications of research on children's learning for standards and assessment: A proposed learning progression for matter and the atomic molecular theory. *Measurement: Interdisciplinary Research and Perspectives*, 4(1&2), 1-98.

Staley, D. (2002). A History of the Future. *History and Theory*, 41, 72-89.

Staley, D. (2007). *History and future: Using historical thinking to imagine the future*. Lexington Books.

Toulmin, S. (2003). *The Uses of Argument*. USA: Cambridge University Press.

Triola, M. (2009). *Estadística*. México: Pearson Education.

Van Drie, J., & Van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review*, *20*(2), 87-110. doi:10.1007/s10648-007-9056-1

Van Drie, J., Van Boxtel, C., & Van der Linden, J. L. (2006). Historical reasoning in a computer-supported collaborative learning environment. In A. M. O'Donnell, C. E. Hmelo, & G. Erkens (Eds.) Collaborative learning, reasoning and technology (pp. 265–296). Mahwah NJ: Erlbaum.

Vygotsky, L. (1978). *Mind in society*. London: Harvard University Press.

Vygotsky, L. (1986), *Thought and Language*. London: The Massachusetts Institute of Technology.

Voss, J. (2006). Toulmin's Model and The Solving of Ill-Structured Problems, In D. Hitchcock y B. Verheif (eds.) *Arguing on the Toulmin Model: New Essays in Argument Analysis and Evaluation*, 303-311.

Waring, S., & Torrez, C. F. (2010). Using digital primary sources to teach historical perspective to preservice teachers. *Contemporary Issues in Technology and Teacher Education*, *10*(3), 294-308.

Wineburg, S. (2001). *Historical thinking and other unnatural acts. Charting the future of teaching the past.* USA: Temple University.

Wineburg, S. (2007). Unnatural and essential: the nature of historical thinking. *Teaching History*, 129, 6-11.

Yin, R.K. (1994). *Case Study Research – Design and Methods, Applied Social Research Methods* (vol. 5, 2nd ed.). Newbury Park, CA, Sage.

Sergio Méndez is a Professor of Psychology at the Autonomous University of Mexico,

Felipe Tirado is a Research Professor at the Autonomous University of Mexico

Contact Address: Av. Los Barrios No 1, Tlalnepantla, Estado de Mexico, Facultad de Estudios Superiores Iztacala, National Autonomous University of Mexico. 54090. Sergio Méndez' email: serchazubi@gmail.com

Dialogic Readers Children Talking and Thinking Together about Visual Texts

Lirio Flores Moncada[1]

1) University of Deusto, Spain

PLEASE SCROLL DOWN FOR ARTICLE

Review

Maine, F. (2015). *Dialogic Readers Children Talking and Thinking Together about Visual Texts*. London: Routledge.

What is happening when children co-construct meaning while making a dialogue and interacting? In the book, *Dialogic Readers Children Talking and Thinking Together about Visual Texts*, Fiona Maine responds this inquiry and provides a wider clarification in this field. Fiona Maine is a lecturer in literacy education at the University of Cambridge, expert in children's critical and creative thinking as they engage with a range of text modes (written, visual and moving-image). She has also years' experience working as a teacher and with teachers, as they develop their own practice.

First, she plunge us in a reading where the reply arises from seven blended chapters. The result is exposed through an exhaustive and detailed theoretical review on language of co-construction, children's critical and creative response to text, dialogic transaction between text and readers and the use of language as a tool for creating a social cohesion between readers. Furthermore, her study includes an analytical framework that illuminates the dialogic and co-constructive functions of the language used by children as they make meaning together from visual text.

The book is supplemented with case studies, which in addition to clarifying and exemplifying the theoretical arguments, give meaning to the whole explanation. It should be stressed that throughout the text, educational notions and practices of how to implement them at the classroom level primary aim are proposed. It remains to say that the author argues how important is performing reading through visual and moving images as they are essential in this 21st century notions of literacy.

In the core of this text we find the reason why dialogue through moving images is so important. On one hand, she poses that the solution to this problem helps to develop children's criticism abilities and convergent thinking, as they are free to express their own thoughts, reasons, hypotheses, ideas and/or results about visual text. So these thoughts are confronted and

evaluated by their interlocutor. On the other hand, she proposes the story spinning, being here where the creative and divergent process is carried out itself. Creative, because each participant explains and argues what is happening from his own social and cultural context. And divergent, because on theses meaning-making processes, agreement is not necessary.

To evaluate processes that occur in these two modes of thinking, the author rightly takes up bases provided by the Vygostskian and Piagetian theories, which she applies objectively in a consistent way, as they are indispensable to understand dialogic reading.

Maine reminds us in an informed manner how thought is produced from social interaction as well as the potential for cognitive development depends on the "Zone of proximal development" (Vygotsky, 1978).

Likewise on the hypotheses, reasons, explanations and resolutions given by children in their visual text reading, Piaget's processes of accommodation and assimilation are clearly identified (Piaget, 1952) . Extracts as the following dialogue: "02 Gina: Yeah…there's lots of men and they kid of [seem to be]…03 Sophie: [Reminds me of] Mary Poppins…04 Gina: Hmm, but a male version…." are a good example of these processes.

The implications of the contributions offered by the book to educational context, specifically for classroom practice, are critical and develop in children cognitive processes as critical thinking, creativity, responsibility, collaborative and reflective spirit, among many others skills.

The book promotes values for future generations. It teaches how to dialogue and respect different ideas, hypotheses, and opinions. It promotes conflict resolution and teaches how to listen and how to reach the right to be heard in freedom. It must be taken into account that dialogue opens a possibility for improvement in any context, especially in education. It is important to note that the use of visual texts (images, moving images and picture books) reduces barriers to dialogue that arise due to misunderstanding of words and written texts.

Finally, I found inspiration in Maine's words, which personally shocked me and pushed me to keep reading her book: "[…] we should stop and listen carefully what children say".

References

Maine, F. (2015). *Dialogic Readers Children Talking and Thinking Together about Visual Texts*. London: Routledge.

Piaget, J. (1952). *The origins of intelligence in children*. International Universities Press, Inc.

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

Lirio Flores Moncada
Faculty of Psychology & Education
University of Deusto
lirioflores@deusto.es