Articles

ATHENS INSTITUTE FOR EDUCATION AND RESEARCH
A World Association of Academics and Researchers
8 Valaoritou Str., Kolonaki, 10671 Athens, Greece.
Tel.: 210-36.34.210 Fax: 210-36.34.209
Email: info@atiner.gr URL: www.atiner.gr

(ATINER)   *Established in 1995*   *(ATINER)*

Mission

ATINER is a *World Non-Profit Association* of Academics and Researchers based in Athens. ATINER is an independent **Association** with a **Mission** to become a forum where Academics and Researchers from all over the world can meet in Athens, exchange ideas on their research and discuss future developments in their disciplines, **as well as engage with professionals from other fields**. Athens was chosen because of its long history of academic gatherings, which go back thousands of years to *Plato's Academy* and *Aristotle's Lyceum.* Both these historic places are within walking distance from ATINER's downtown offices. Since antiquity, Athens was an open city. In the words of Pericles, **Athens"…*is open to the world, we never expel a foreigner from learning or seeing".* (**"Pericles' Funeral Oration", in Thucydides, *The History of the Peloponnesian War). It is* ATINER's **mission** to revive the glory of Ancient Athens by inviting the World Academic Community to the city, to learn from each other in an environment of freedom and respect for other people's opinions and beliefs. After all, the free expression of one's opinion formed the basis for the development of democracy, and Athens was its cradle. As it turned out, the Golden Age of Athens was in fact, the Golden Age of the Western Civilization. *Education* and *(Re)searching* for the 'truth' are the pillars of any free (democratic) society. This is the reason why *Education* and *Research* are the two core words in ATINER's name.

Athens Journal of Education
Editorial and Reviewers' Board

President's Message

All ATINER's publications including the e-journals are open access without any costs (submission, processing, publishing, open access paid by authors, open access paid by readers etc) and are independent of the presentations made at any of the many small events (conferences, symposiums, forums, colloquiums, courses, roundtable discussions) organized by ATINER throughout the year. The intellectual property rights of the submitted papers remain with the author.

Before you submit, please make sure your paper meets some [basic academic standards](#), which include proper English. Some articles will be selected from the numerous papers that have been presented at the various annual international academic conferences organized by the different [divisions and units](#) of the Athens Institute for Education and Research.

The plethora of papers presented every year will enable the editorial board of each journal to select the best ones, and in so doing, to produce a quality academic journal. In addition to papers presented, ATINER encourages the independent submission of papers to be evaluated for publication.

The current issue of the Athens Journal of Education (AJE) is the second issue of the sixth volume (2019). The reader will notice some changes compared with the previous issues, which I hope is an improvement.

Gregory T. Papanikos, President
Athens Institute for Education and Research

Athens Institute for Education and Research
A World Association of Academics and Researchers

21st Annual International Conference on Education
20-23 May 2019, Athens, Greece

The Education Unit of the Athens Institute for Education and Research (ATINER) organizes its **21st Annual International Conference on Education, 20-23 May 2019, Athens, Greece** sponsored by the Athens Journal of Education. The aim of the conference is to bring together scholars and students of education and other related disciplines. You may participate as **stream leader**, presenter of one paper, chair a session or observer. Papers (in English) from all areas of education are welcome. Please submit a proposal using the form available (https://www.atiner.gr/2019/FORM-EDU.doc).

Academic Members Responsible for the Conference

- **Dr. Alexander Makedon**, Head, Education Research Unit, ATINER.
- **Dr. Mary Ellis**, Director, Human Development Division, ATINER & Senior Lecturer, National Institute of Education, Singapore.

Important Dates

- Abstract Submission: **8 April 2019**
- Acceptance of Abstract: 4 Weeks after Submission
- Submission of Paper: **22 April 2019**

Social and Educational Program

The Social Program Emphasizes the Educational Aspect of the Academic Meetings of Atiner.

- Greek Night Entertainment (This is the official dinner of the conference)
- Athens Sightseeing: Old and New-An Educational Urban Walk
- Social Dinner
- Mycenae Visit
- Exploration of the Aegean Islands
- Delphi Visit
- Ancient Corinth and Cape Sounion

Conference Fees

Conference fees vary from 400€ to 2000€
Details can be found at: https://www.atiner.gr/2019fees

Athens Institute for Education and Research
A World Association of Academics and Researchers

3rd Annual International Symposium on "Higher Education in a Global World", 8-11 July 2019, Athens, Greece

The Education Unit of ATINER is organizing the **3rd Annual International Symposium on "Higher Education in a Global World", 8-11 July 2019, Athens, Greece** sponsored by the Athens Journal of Education. The aim of the symposium is to examine educational developments throughout the world in universities, polytechnics, colleges, and vocational and education institutions. Academics and researchers from all areas of education are welcomed. You may participate as stream organizer, presenter of one paper, chair a session or observer. Please submit a proposal using the form available (https://www.atiner.gr/2019/FORM-COLEDU.doc).

Important Dates
- Abstract Submission: **11 March 2019**
- Acceptance of Abstract: 4 Weeks after Submission
- Submission of Paper: **10 June 2019**

Academic Member Responsible for the Conference
- Dr. Sharon Claire Bolton, Professor, Management Unit, ATINER & Dean, The Management School, University of Stirling, Scotland.
- Dr. Mary Ellis, Director, Human Development Division & Senior Lecturer, National Institute of Education, Singapore.
- Dr. George Priovolos, Director, Center for Small and Medium-Sized Enterprises (CSME) & Professor, Iona College, USA.

Social and Educational Program
The Social Program Emphasizes the Educational Aspect of the Academic Meetings of Atiner.
- Greek Night Entertainment (This is the official dinner of the conference)
- Athens Sightseeing: Old and New-An Educational Urban Walk
- Social Dinner
- Mycenae Visit
- Exploration of the Aegean Islands
- Delphi Visit
- Ancient Corinth and Cape Sounion
    More information can be found here: https://www.atiner.gr/social-program

Conference Fees
Conference fees vary from 400€ to 2000€
Details can be found at: https://www.atiner.gr/2019fees

Requisite Community Engagement for Teacher Education: A Different Take on Service Learning

By Deborah Biss Keller[*]

This article examines pre-service teachers' responses to their experiences with community engagement as a service learning project as part of an introductory Education course at a large, midwestern urban university in the United States. It is typical for students in pre-service Education courses to participate in service learning in schools or community centers, but it is less common for Education courses to require service learning projects in which the focus is for students to engage with adult community members. I argue that it is imperative for students to become familiar with communities in which they teach and learn from residents' "funds of knowledge" (Gonzalez, Moll, & Amanti, 2005). A service learning project with a community center, the focus of which is empowerment of local residents, is described. Located in an urban neighborhood in which most of the residents are of a racial minority different from most of the pre-service students themselves, the center provided the space for the class to learn from adult community residents about issues surrounding empowerment and community engagement. Drawing upon Freire's (2003) notion of praxis I present findings of a qualitative analysis of students' response papers in the context of a curriculum that focuses on critical social justice and the development of culturally competent teachers.

Keywords: Agency, Community engagement, Critical social justice, Pre-service teacher education, Service learning

Introduction

Service learning for teacher education has become part of the curriculum in many pre-service teacher programs. While Education course work often includes students participating in asset based studies of school children's/youths' communities, service learning typically takes the form of students working with children/youth in schools or community centers and reflecting about the experience with regard to one's own biases, pedagogical techniques, and cultural competency among other elements. What is not as common, however, is service learning that engages students working with adults in the community in order to learn more about the community and residents in general and the residents' funds of knowledge, defined by Gonzalez, Moll, & Amanti (2005) as encompassing the idea that "people are competent, they have knowledge, and their life experiences have given them that knowledge" (preface, pp. ix-x). According to them, this concept undergirds the depiction of communities with regard to their resources and how to utilize them in the classroom. Their aim is to provide a lens in which households in "working-class or poor communities" (preface, p. x) are seen

---
*Clinical Associate Professor, Indiana University-Purdue University, USA.

predominantly through their strengths and their assets, which can inform pedagogical practice. Similarly, Buck & Sylvester (2005), in a project with Masters students in Elementary Education at the University of Pennsylvania, described the basic goal of having their students do research on urban neighborhood communities as viewing "urban families and communities as resources of talent, knowledge, and possibility" (p. 214). They noted that such research and curriculum can facilitate an environment for social change and student agency. This is in direct contrast to deficit theories, which hold that minority, low-income children do poorly in school because their backgrounds are "linguistically or cognitively" lacking (Collins, 1988, p. 300), or that they are culturally deficient (Payne, 2005, as cited in Sensoy & DiAngelo, 2017, p. 82). This fails to acknowledge systemic and structural forces at work (Collins, 1988; Sensoy & DiAngelo, 2017), including inequities in education and "basic human rights" (Gorski, 2008). Dewey (1932/1990) contended that school and community should be directly connected, and he called for collaborative work among students and asserted that the "helping" of another, "instead of being a form of charity which impoverishes the recipient, is simply an aid in setting free the powers…of the one helped" (p. 16). According to Freire (1998), a teacher should "discuss with students the concrete reality of their lives" (p. 36), and he questions why one should not create an "'intimate' connection" between general curriculum and "knowledge that is the fruit of the lived experience of these students as individuals" (p. 36). Ladson-Billings (2006) wrote that "we lack of complex understandings of how individual, family, community, school, and societal factors interact to create school failure for some students" (p. 106). In questioning a group of teachers regarding whether their school staff had attended events in the community in which they taught, she found that none of them had done so. She noted how many preservice teachers do not see themselves as having a culture due to their belonging to the dominant culture, which has become normalized. She stressed the need for students to become familiar with the culture of communities where they will teach and examine how their own culture informs their beliefs and practices. In doing so, she maintained, "they may become more open to the power of culture to shape the learning and experiences" (p. 109) of their future students. In their work with Native American students, McCarty and Lee (2014) found striking a balance between "academic, linguist, and cultural interests" necessitates "direct accountability to Indigenous communities" (p. 119). Accountability to communities for the education of their children is paramount for effective, culturally sustaining pedagogy, and drawing upon communities' funds of knowledge is integral to this endeavor.

Here I describe a service learning project in which students in an introductory education class at a large, midwestern urban university in the United States did community work and engaged in dialogue with adult residents from urban neighborhoods in an effort to help students understand the importance of funds of knowledge that children and youth bring with them from their communities into the classroom. Due to the qualitative nature of the study and the examination of student responses, I looked for overall themes that emerged from their responses. Therefore, the broad questions I wanted to address were: 1. How did engagement

with adult residents from an urban community affect students' learning and attitudes about the importance of spending time getting to know the community in which one teaches? 2. How did engagement with adult residents from an urban community influence students' perceptions of the importance to them as prospective teachers of PK-12 students' and their families' "funds of knowledge"?

Participants and Context

The 23 participants were students enrolled in a Themed Learning Community (TLC) that included my class, an introductory Education class; a History class; a Writing class; and a First Year Seminar class. Students attended all of these classes together. The theme revolved around social justice, and the First Year Seminar was part of a summer program so therefore did not meet every week during the regular semester. The time slots for the First Year Seminar and my class were back-to-back, allowing us to combine the two class times for the students and myself to go to the service learning site and participate as a group. The other instructors of the TLC participated as they were able. The service learning project was with a local urban community center that focuses on empowerment of youth and other community residents and is located in a primarily African American neighborhood. Each service learning session was two hours, and the initial plan was to have two sessions that involved group dialogue and two sessions that involved gathering soil samples in city neighborhoods to be tested for lead. My students would just do the collecting of samples, not the testing. This was part of a larger project funded with grant money for a scientist who was using the grant to work with the community center. Due to a logistical issue, with those involved in granting permission to proceed with the collection of soil samples, we were unfortunately unable to collect the samples, as the permission was not granted during our service learning time spent there. When we began the sessions it was not known that we would not be able to do the collecting; therefore, the first two hours consisted of a dialogue and information session in which the scientist involved with the project talked about the adverse effects of lead poisoning on children. He involved my students in discussing possible solutions and their own roles as prospective teachers in working with educational interventions with students who are affected by lead poisoning. Students learned about environmental racism and its ramifications, how some people are affected more than others due to their race because of lack of soil testing and clean-up where they reside.

By the second session permission to collect soil samples had still not been granted, so the class raked leaves at a cemetery near the center, and the leaves would be used as compost for the center's urban garden. Prior to going to the cemetery students learned from the leaders of the center about the importance of social networking and social capital and how these can provide opportunities for residents to engage in entrepreneurship. The leaders explained how the raking would provide a service both to the cemetery and to the center, and consequently to the neighborhood residents.

The third session involved a dialogue with adult community members about

the history of racism and how racism still plays out in schools and what my students as prospective teachers, most of whom were white, could do to become more culturally competent and more effectively work with students of color. Students were able to listen to voices of those directly affected by racism past and present. During this discussion topics such as lack of cultural competence among white teachers in an integrated school, lack of voice from People of Color about integration, lack of protection of youth of color in integrated settings, the importance of listening to People of Color, the importance of knowing one's own culture before trying to understand others', the importance of keeping a check on one's own assumptions, the importance of compassion, and the necessity of increasing resources in urban spaces were addressed, among others.

By the time of the last session we still had not been given permission to collect soil samples, so we went to the center's aquaponics garden and learned about how it works and serves the community. This was followed by a gathering at the center to view a video about a well-known activist.

When initially working with the community center to seek an opportunity for my students to become involved in the community to challenge deficit mindsets and appreciate residents' funds of knowledge, I left it up to the directors of the agency to determine what we would do, though I did request a dialogue with community members. I wanted my students to hear the voices that often get left out of curriculum and to examine their (my students') assumptions and beliefs about those from groups different than their own. This self-reflection aligns with our School of Education's and our particular TLC's focus on critical social justice and anti-oppression pedagogy. Typically I have sent students to urban schools or community centers to work with pre-kindergarten through twelfth-grade students, or a local program that works with students aged 18-21 who have disabilities. This work has involved tutoring and/or working in some other capacity with children/youth. That is what most if not all of the students expected for this service learning project. Therefore, there was some resistance to taking part in a project such as the one in which we were going to engage. I heard through another instructor in the TLC and from students' written reflections of the frustration of some students that they were not going to work with children/youth. I had anticipated this and constructed reflection prompts accordingly, asking students to specifically link what we were doing to their future roles as teachers. These reflective writings provide the data for this project.

Theoretical Framework

I employed the concept of Funds of Knowledge (Gonzalez, Moll, & Amanti, 2005) as described above and Freire's (2003) notion of *praxis* – "the action and reflection of men and women upon their world in order to transform it" (p. 79) to inform my selection of the service learning project, the reflection prompts, and the analysis of the reflections. Freire contrasted the "banking approach" of education, in which "the students are the depositories and the teacher is the depositor" (p. 72), with the "problem-posing" approach, in which students and teacher construct and

negotiate knowledge together through dialogue. In the latter approach the roles of the teacher and students take on a reciprocity; each becomes the teacher *and* the student – the learning involves "teacher-student with students-teachers" (p. 80). For Freire, this was crucial for emancipatory pedagogy, that which he referred to as *engaged pedagogy*; the teacher engages with the students in a quest to uncover reality as it exists in lived experience. Freire asserted that "education as the practice of freedom—as opposed to education as the practice of domination— …denies that the world exists as a reality apart from people" (p. 81). It was in this spirit that I chose to have my students focus on learning about the community through engagement with a center the aim of which is to empower neighborhood residents for leadership and sustainability. Students did not just learn about the neighborhood's funds of knowledge; they also learned from community members themselves of the challenges past and present that People of Color face due to systemic racism. Although my class addressed oppression in its many forms and how it impacts the lives of children and youth, coming from urban community residents as opposed to coming from their White instructor provides a more authentic, credible voice. This of course assumes that the students will listen to these voices and see them as valid. The hope is that with curriculum that focuses on critical social justice prior to, during, and following the service learning sessions, including reflective writing that requires that the students at least consider the purpose of this engagement, they will begin to develop critical consciousness.

Data Collection and Methodology

Data consisted of students' answers to prompts that I gave them for each session. While all 23 students participated, two students did not complete the written assignments for the third and fourth sessions but attended both; one student did not complete the written assignment for the fourth session but attended; and one student did not complete the written assignment for the first session but attended and did not attend the third and fourth sessions. Each written assignment was to be 800 words. Students' answers to the prompts were qualitatively examined for general emergent themes. Responses to some question sets overlapped with those of other sets. Therefore, I looked for overall themes from the reflective writings. When a response fit under more than one theme I selected a thematic category for the sake of analysis; an exception is the category of *Themes from Students* – due to the students' listing themes from their service learning, some of those themes overlapped with general overall themes that emerged in the study, and they are included in both categories. To insure credibility in the qualitative analysis I present students' words verbatim in excerpts from the reflection responses. While I necessarily selected particular excerpts to include here, I note the frequency of responses, as relayed in numbers of students who gave similar or like responses. In addition, if it was not clear that a student's response fit in a category I did not include it. For example, if a student mentioned "funds of knowledge" in a response but it was not elaborated upon or it was clear

that the student did not understood the significance of the term, then I did not include it.

The first question set asked students to consider how the discussion on lead poisoning enlightened them as future teachers and how it relates to oppression, awareness, and critical social justice. They were also asked to relate the discussion to agency/empowerment and explain why this is important. The second question set prompted students to think about why it is important to learn about residents of the community before going to work in the schools, and what that has to do with children's/youth's and their families' funds of knowledge and working toward hearing their voices, and why this understanding is crucial. Students were asked how this affects their attitudes about those who are from groups different than their own. And, due to the change of service work, they were asked how raking and bagging leaves plays a part in the role of social capital in the projects at the center. The third question set asked students what they learned that they might not have learned in a school environment. They also were asked to explain what they learned about Whiteness, desegregation, assumptions, and student engagement, and they were asked how they can prepare themselves to teach in an urban school. The final question set prompted the students to draw upon our course curriculum, including service learning and other TLC curriculum as appropriate, to consider how aquaponics and the activist about whom they viewed a documentary interconnect and to relate it to their responsibilities as future teachers. In addition, they were asked to think of some general themes they learned through their service learning sessions that will help them become effective teachers, and how their sessions will influence them in this endeavor.

Findings

Main themes that emerged involved ideas around awareness, community, failure to see relevance to the curriculum, empowerment versus savior mentality, social capital, funds of knowledge, things learned that students believed would not have been learned in a school setting, discrimination and Whiteness. I also include a section on overall themes from the students' perspectives.

Importance of Awareness

Awareness was a recurring theme throughout the course, both self-awareness and awareness of how oppression manifests itself at the macro and micro levels. The service learning experience provided students with authentic perspectives that underscored what we were addressing in class. Eighteen students (78.3%) wrote about awareness at the personal level or as it related to the topics addressed in service learning.

One student, speaking of being a teacher in the future and in relation to lead poisoning and other issues that impact education, said:

I wish to actively bring these ideas up in class, or maybe hold discussions outside of class. These discussions could involve students and parents alike; bringing awareness to issues like this is crucial to make progress.

This student expressed the importance of taking the awareness into the classroom and to parents, indicating that broader awareness is necessary to make "progress." This aligns with the center's emphasis on empowerment.

Another student said:

These issues are important because without this knowledge I would have never known that lead poisoning had long-term effects on the environment, children, and the community. Becoming educated on these issues can make or break the communities safety, and the safety of young children.

A student, referring to a lack of action at the structural level to address lead poisoning in a lower-income, minority neighborhood said:

This is an example of environmental racism which was the lack of awareness people made about a specific area because of the minorities that lived there.

Here the student is naming a form of oppression as it relates to the environment.

One student wrote about self-awareness of assumptions:

One thing I really learned from [the center] today was to go into urban schools with no prior assumptions. I think this is something a lot of people don't think about. I have a lot of assumptions about urban schools that I never realized really existed in my mind until I started taking this class.

Additionally, a couple of students addressed *lack of* awareness of one's own oppressive behavior. One student, pointing to the fact that lead poisoning isn't addressed or prevented due to its affecting low-income communities of People of Color, said:

This falls under what we talk about a lot in class and how discrimination happens and maybe people don't even realize it.

Here the student is referring to how oftentimes people with power are not always aware that they are discriminating. Sensoy & DiAngelo (2017) noted that most discrimination is not in our conscious awareness; we are socialized to believe certain ways and internalize these beliefs and assumptions about others. They also asserted, however, that those who reap the benefits of discrimination might not be interested in understanding the effects of it on others. My students were challenged throughout the course to examine their own biases and assumptions and to consider where they come from. One student said:

> We think that just because we don't have an issue, that the [sic] everyone around us don't [sic] have an issue.

Community

Twenty students (87%) related the experience in the community to their roles as future teachers.

One student revealed an understanding that children/youth are significantly influenced by experiences outside of school:

> This experience allowed me to understand that the more I understand my community and my students, the better I will be able to teach. I also now understand that my students are and can be heavily affected by their previous surroundings outside of the classroom.

Another student exhibited a deeper level of critical thought:

> This is the most crucial component of community: working together for the common good. We live in a hierarchical society, but we must come to realize through our work as educators, citizen scientists, and members of the community, that we are one. We live in an interconnected society, and the actions either taken or untaken by one, affect us all.

Similarly, another said:

> Teaching is a community-oriented job....that is why it is going to be crucial for me to know the community I am in before I even begin teaching there….Students do not live in the classroom, and what goes on before, after, and on days off from school is just as important to their learning and development as what occurs in the classroom.

Another student suggested that "having roots" in the community helps develop trust with parents:

> If the teacher has roots in the community I think it makes the parents more at ease when handing their children over to you. They will know you better and trust you and your judgement if you have had a good influence on the community.

This is parallel with what Freire (2003) had to say about the significance of dialogue to trust. He maintained that in order to secure trust one's actions must match one's words.

One student included their role in the context of both the community and of an educator, saying:

We as a community will need to see to the change, and as a future educator, it is important that I continue to teach this material to more than just those enrolled in classes.

Freire's (2003) notion of reciprocity is reflected in another student's comment:

Educating myself and learning just as much from the students as they do from me is key.

Failure to See Relevance to the Curriculum

As noted above, some students did not expect to dialogue with adults or rake leaves. They expected that they would be working with children/youth. While some students indicated that they initially did not see the relevance in what they were doing for service learning but that they eventually did understand the purpose, a few asserted their frustration at what they perceived to be irrelevant to the curriculum.

One student said:

The first session didn't really relate to the material we are learning because we just talked about soil sampling and the effects that lead poisoning has on a person. Hopefully the next sessions will relate to what were [sic] are learning more than the first.

Another said:

During this service trip we went to a cemetery and racked [sic] leaves. We put the leaves in a trash bag so they could use them to compost. I don't know how this relates to education or has anything to do with it…I could see if we took our students out to help clean up the city but that is different than what we did. I think we should be doing other things to help us as future educators.

It is clear that these students did not make the connections between teaching and the importance of working *with* the communities where they teach to empower the residents, some of whom could be their future students' families. This lack of connections made is despite the fact that links were made clear to the students by the people involved at the center. It is difficult to ascertain whether the students genuinely did not understand or they did not wish to understand. Teaching social justice typically brings resistance, albeit with some classes and students more than others.

Empowerment versus Savior Mentality

Nine students (39.1%) demonstrated an understanding of the dangers of a

savior mentality.

One student said:

This service learning project allows for a community to collectively promote change, rather than an outsider like [the scientist] to be the "savior" and "help" the neighborhood.

This comment reflects the student's connecting the lead testing project to collaborative efforts.

Another student noted a self-check on language used:

When I first typed the second main point I typed "that it is important to help" but I decided to change to "get involved" because, as we talked about in class, then there is less likely to have the savior mentality.

Another student applied the idea of agency to children/youth:

If a child is struggling, we should ask them what they need or are going through, instead of insisting our [sic] own solution. We want to encourage the students to seek out their own change, be it in their education, groups they are a part of, or even at home.

One student wrote about using White privilege to help empower others:

If you have a false view of "good neighborhoods" and "bad neighborhoods" is [sic] is easy to fall prey to the savior mindset…Through the societal and institutional power I hold as a white person, I am responsible for listening to their desires and aiding them in being heard. The power I hold in society is undeniable. Though it is unjustly granted, it can be used to aid in empowering those whose voices are chosen not to be heard by those in the institutions.

These reflections align with what Freire (2003) had to say regarding attempts to liberate those who are oppressed. He maintained such attempts must only be made with the oppressed's "reflective participation in the act of liberation" (p. 65), for to do otherwise "is to treat them as objects which must be saved from a burning building…it is to transform them into masses which can be manipulated" (p. 65). Similarly, in his call for problem-posing education, he stated, "The teacher's thinking is authenticated only by the authenticity of the students' thinking. The teacher cannot think for her students, nor can she impose her thought on them" (p. 77). When engaging in critical pedagogy, teachers must work *with* the students and they and the teacher in turn must work *with* the community. For Freire, dialogue is crucial for transformation; one cannot speak for another in a way that denies the other their voice.

Social Capital and Funds of Knowledge

While students seemed to overall demonstrate an understanding of social capital, only just over half (52.2%) exhibited a clear understanding of "funds of knowledge."

A student described the relationship the center has with other entities in the community:

> In exchange for our work, [the center] got to keep leaves for composting…. [The center has] formed a relationship where the restaurant relies on them weekly to pick up their scraps.

Another student connected funds of knowledge with social capital by placing "us" in the role of facilitating collaboration among communities, though it is not clear what the student meant by "because we know that these are people who will continue to work collaboratively *within* [emphasis added] the community":

> When we encourage agency within a community, we are bringing new people in. These people have their own individual funds of knowledge to contribute to the community. By sharing these connections with people across many communities, we foster a strong foundation of social capital, because we know that these are people who will continue to work collaboratively within the community.

The following excerpt illustrates a degree of understanding of the importance of funds of knowledge; however, it is unclear what the student was suggesting when referring to "developing the students' and families' funds of knowledge." Perhaps by "developing" the student meant "utilizing," as the student went on to describe the importance of incorporating funds of knowledge into the curriculum.

> Building that relationship with my students and their families and then bringing their culture into the classroom and digging into their prior knowledge is going to be crucial in developing the students' and families' funds of knowledge. What I learn about my students and then how I incorporate that into meaningful, culturally correct and positive lessons in the classroom is possibly the most important future tool I will need.

Things learned that Students Believed would not have been learned in a School Setting

Eleven students (47.8%) wrote about specific things they learned from the service learning project that they felt they would not have necessarily learned in a school setting. This was in response to a question prompt intended to have them reflect what is learned in service learning such as what they had versus working with PK-12 students. Some of the responses involved what they learned that they

believed they would not learn as classroom students themselves. Nevertheless, what they had to say about the value of hearing personal narratives is significant to the learning process overall.

One student said:

It made it more personal because these people were here telling their stories…One of them talked about discrimination at [a parochial school in one of the surrounding areas], and I live close by there, and have had countless friends go there. It really caught my attention to realize even very religious people discriminated against African Americans because they were different.

Another said:

I think it was good for us to hear how to prepare for urban schools from people who have a lot of experience in urban communities and school systems. It really shocked me to hear how badly some students are treated in schools today based on their color.
…much of what I learned could not be taught or absorbed in a regular school setting because it involved real-life people and their personal stories and experiences…

One student's reflection conveyed how the experience complemented course curriculum:

During this session at [the center], I have been exposed to lives of real people, real moments, and how their journey has had an impact on historical contexts like we learn in class.

Discrimination and Whiteness

Sixteen students (69.6%) demonstrated at least some understanding of the role of discrimination in White privilege. Some comments include:

Since I am white I will never be able to understand what African Americans…go through on a daily basis….I shouldn't act like I know how they feel because I don't and never will, but instead I just need to listen and be a supporter.

The thing about whiteness is, sometimes you do not learn something new but you are retold or reminded of something you have already heard, but chose to very simply ignore. Even if you mean no harm by it, which many of us do not, whiteness is a very real privilege.

My whiteness was never something I thought about previous to our third session at [the center]…Thanks to our Themed Learning Community, as well

as the service learning trips, I now know why; my life has never been challenged because of the fact that I am white….Whiteness in general grants those people the safety and success that people of color have to work so hard for…

I learned that is [sic] very easy for me to forget how privileged I am as a white person. As I was listening to all of the stories people from the community provided I felt upset because I have never taken the time to think about these real life issues that people around me are experiencing.

…as white people, we can choose to flee an uncomfortable situation and we don't have to engage in conversations that are uneasy for us. People of color don't have this privilege.

Before learning about my whiteness I only saw things through my white lense [sic] of perception. When we do not look at all sides and reeducate ourselves it ignores the contributions of POC [People of Color]….POC are not here to help me work with my whiteness…Something that was prevalent in my school was color blindness. "I do not see color." This five word sentence is so oppressive. It denies culture and most importantly it denies oppression.

These comments demonstrate a significant degree of self-reflection. During the course of the semester we discussed race many times in class, and the dialogue with community residents and listening to their stories made it "real" to my students.

The following excerpt addresses the danger of using the bad/good binary when referring to neighborhoods:

…what may be "bad" to us, another student could live with every day. For instance, when referencing neighborhoods that seem different than the one that whites may have been privileged enough to live in, we want to avoid saying it is bad because that could be one of your student's neighborhoods.

Themes from Students

In the final question set students were asked what some of the main themes were that they learned through their service learning that will help them become effective teachers. Fifteen students (65.2%) specifically addressed themes. Their answers varied, but most conveyed a sense of new knowledge and the importance of involvement. Seven students wrote of community involvement; four wrote of the importance of community in general; three wrote of discrimination or oppression; three wrote of building relationships. Other less-often mentioned themes were social justice in general, education/learning, kindness, supporting others, public awareness, debunking stereotypes, importance of social capital, empowerment, inclusivity, culture, diversity, respect, and listening. It is important

to note that some of the themes overlap. For example, public awareness can involve community involvement and supporting others, and kindness can involve respect and listening, and social justice can potentially relate to all of the themes, etc.

One student expressed disappointment at not being able to do more service but noted things learned through the experiences at the center:

> At [the center], the most important lesson that has carried me through the sessions is to get involved. Don't just expect change to happen, but actually do something about it. It is disappointing…that only service [sic] we were able to do was rake leaves. I wish we could have really gotten into a community, and see what it's like to make a change and be involved. Another lesson is the theme of all our classes, of social justice. I learned how deprived some people are of resources, and how unfair it is. Social justice really spoke to me, when I was at [the center] and all the different people from the community came to tell their story[ies]…

Interestingly, this student conveyed a wish that service learning had involved getting "into a community" to experience making a change but also noted how social justice "spoke" to her when the community residents shared their narratives. The student seemingly did not view raking leaves for the center's composting efforts for the community garden as involvement.

Another student wrote about learning about social justice from different perspectives:

> I've really enjoyed getting different perspectives about social justice and what that means to different people. I think [the center] especially helps me get to see these different views because I come from a small town and I not [sic] used to big cities and diversity. I think [the center] has helped me experience some of this and really get a good idea of what social justice is and what I can do to help the community as a teacher of the future generations.

Another student wrote about feeling more confident to teach and supporting others in their work for others' rights:

> After working with [the center], I feel more motivated in [sic] confident in my ability to reach out and want to be a part of something bigger. I also feel a lot more confident in my ability as a teacher to motivate and inform others on the importance of standing up for what they believe in, as well as supporting those who are standing up for the rights of others.

One student wrote about seeing beyond stereotypes and recognizing the role of systemic issues in people's lived experiences:

I used to think that some of these stereotypes were real until I went out into the community. I saw that people in these conditions really want to see change. There are people out there working two jobs and still can't get out of their situation. It is like this because the system is messed up. That is why it is up to people in the dominant group to see that there is a problem and we need to use our voice to help them in a way that they need us too [sic].

Another student wrote about learning of the importance of becoming involved:

I would have never thought, or at least thought about in great depth how important it is to be involved… [The center] has taught me that teaching is far beyond the classroom.

One student expressed how the service learning experience highlighted the importance of being aware of her Whiteness:

During this service learning experience, I noticed that it was very easy for me to forget about my white privilege before I started getting involved with the urban community. As a white person I don't have to think about the majority of the things I do throughout my daily life. I also don't need to think about the other cultures that surround me because I am in the majority. However, after this experience I realized how important it is to be aware of my whiteness.

A student wrote about how service learning inspired her as a Person of Color to share her own experiences with her future students:

The main theme I learned from [the center] is that building relationships is really important to education. Racism is always going to be there. I'm [a Person of Color] and I will be teaching a variety of students as well, so telling them from personal experience and being that person they can always go to is what I learned from [the center].

Another student expressed the importance of learning how to listen:

Before I can be the voice for others, I have to first learn how to listen. Listen to the community *and* listen to what is needed from me without telling them what they need.

Discussion

Most students demonstrated the need for teachers to be involved in the community in which they teach so they would be more culturally competent and effective teachers. This aligns with Paris' (2012) call for a *culturally sustaining pedagogy* in which students are supported "in sustaining the cultural and linguistic competence of their communities while simultaneously offering access to

dominant cultural competence" (p. 95). Many students also expressed the need to listen to residents to hear what they needed, versus going in with a savior mentality, and work with the residents of a community to bring about needed changes. This is parallel with Freire's (2003) notion of *praxis*, "the action and reflection of men and women upon their world in order to transform it" (p. 79). There was evidence in students' written responses that overall they understood the significance of projects such as testing soil for lead and its direct link to teaching children/youth who are affected by lead poisoning. This information provided students with an opportunity to better understand environmental racism and its effects. Students overwhelmingly expressed the value of hearing residents' narratives about racism and its manifestations in educational spaces. This again speaks to the importance of communities' "funds of knowledge" (Gonzalez, Moll, & Amanti, 2005). While some did not view raking leaves as relevant to the curriculum, many saw the relevance of the overall service learning project and wrote about it in terms of working with the community and the importance of this as it relates to their future students.

Conclusions, Implications, and Suggestions

While the project did not go as initially planned, this in itself afforded the students a learning opportunity for understanding the need to be flexible and for drawing connections to education in ways they might not otherwise have considered. The experience proved to be valuable on numerous levels. Students were able to hear the voices of the disenfranchised and see the possibilities for empowering their own students and their families. The experience provided students with a "big picture" approach to learning about lived experiences of others: they were able to reflect upon the inequitable allocation of resources and how this directly influences children's/youth's opportunities for education; they were forced to consider the importance of social networking and agency; they saw how a community center works with residents for empowerment, sustainability, and leadership, and they took part in a small service project that was part of a chain of events that contribute to this endeavor; and White students were forced to grapple with their privilege in a different space, an experience that reinforced our course work. The project proved to be a very different one than the students expected going into the course but one that gave them a wealth of knowledge so imperative to being effective teachers.

Suggestions for future endeavors to provide students with opportunities to learn the importance of knowing the community in which one teaches and its funds of knowledge include the following:

1. Partnering with a community organization for more sessions so that more service can be done, although this can be difficult due to time; class sessions are often not long enough to allow for students and faculty to travel off campus, spend time at the community center, and travel back to campus in time for students' next classes. In addition, finding an extended

time when all students can go at the same time for multiple times outside of class can obviously be difficult or impossible.

2. Have an alternative collaboratively planned in advance with the partnering entity in the case when outside work is part of or all of the service. While things were worked out with our situation such that students ended up having valuable learning opportunities, it might not always be the case that partnering hosts would always be able to accommodate unforeseen changes to planned activities.

This study has significant implications for service learning for students who are planning to be teachers. Working with adult residents in the surrounding community lends itself to teachers' becoming more adept in cultural understanding and acquiring an asset-based mindset as opposed to a deficit one. This involves understanding students' funds of knowledge and embracing their cultural backgrounds and building relationships with parents and community leaders. This is crucial if school children/youth are to be treated equitably and afforded the opportunity to maximize their potentials in a caring, trusting environment.

Implications for further research would include a longitudinal study to follow up with students who stay in a program to examine how what they learn from adult community residents influences their behavior in the field as they move through their Education courses and after the become employed as teachers. Additionally, it would be valuable to the discipline to create and execute a study that places introductory Education students with a community center for a longer period of time per semester and one that is able to plan ahead for unforeseen changes to schedules, weather, etc. Ideally, it would be advantageous to both the community and the students if the partnership could be sustained over more than one semester. Generalizability to other disciplines is possible with areas that involve working closely with children/youth in other community entities.

References

Buck, P., & Sylvester, P. S. (2005). Preservice Teachers enter Urban Communities: Coupling Funds of Knowledge Research and Critical Pedagogy in Teacher Education. In N. Gonzalez, L. C., Moll, & C. Amanti, (Eds.), *Funds of Knowledge: Theorizing Practices in Households, Communities, and Classrooms* (pp. 213-232). Mahwah, NJ: Lawrence Erlbaum.

Collins, J. (1988). Language and Class in Minority Education. *Anthropology & Education Quarterly, 19*(4), 299-236.

Dewey, J. (1990). *The School and Society.* Chicago, IL: The University of Chicago Press. (Original work published in 1932).

Freire, P. (1998). *Pedagogy of Freedom: Ethics, Democracy, and Civic Courage.* (P. Clarke, Trans.). Lanham, MD: Rowman & Littlefield Publishers, Inc.

Freire, P. (2003). *Pedagogy of the Oppressed* (30th anniversary ed.). (M. B. Ramos, Trans.). New York: Continuum.

Gonzalez, N., Moll, L. C., & Amanti, C. (Eds.). (2005). *Funds of Knowledge: Theorizing*

Practices Households, Communities, and Classrooms. Mahwah, NJ: Lawrence Earlbaum Associates.

Gorski, P. (2008). The Myth of the Culture of Poverty. *Educational Leadership, 65*(7), 32-36. Retrieved from https://bit.ly/1bXiWrr.

Ladson-Billings, G. (2006). It's not the Culture of Poverty, it's the Poverty of Culture: The Problem with Teacher Education. *Anthropology & Education Quarterly, 37*(2), 104-109. https://doi.org/10.1525/aeq.2006.37.2.104.

McCarty, T. L., & Lee, T. (2014). Critical Culturally Sustaining/Revitalizing Pedagogy and Indigenous Education Sovereignty. *Harvard Educational Review, 84*(1), 101-125. Retrieved from https://bit.ly/2Bu1PxP.

Paris, D. (2012). Culturally Sustaining Pedagogy: A Needed Change in Stance, Terminology, and Practice. *Educational Researcher, (41)*2, 93-97. https://doi.org/10.3102/0013189X12441244.

Payne, R. (2005). *A Framework for Understanding Poverty.* Highlands, TX: Aha! Process.

Sensoy, Ö., & DiAngelo, R. (2017). *Is Everyone Really Equal? An Introduction to Key Concepts in Social Justice Education.* New York, NY: Teachers College Press.

Opportunities, Challenges and Creative Teaching Practices of New International and Local Recruits in Hong Kong

By Paula Hodgson[*]

Internationalization in higher education in Hong Kong has brought many enriched learning opportunities for students, because international recruits create an alternative learning culture that embraces global perspectives and international research to the curriculum. However, those new to teaching and with non-Asian backgrounds may experience a steeper learning curve as they adjust to a new culture, systems and expectations while working in Hong Kong. Of about 80% local students, educators need to find creative strategies to facilitate them to have interactions and collaborations with non-local students, 15% of whom are from mainland China and have different learning habits. In addition, they have to manage teaching undergraduate and postgraduate programs, which have students from diverse knowledge and working experiences. Twenty-five recruits from social sciences and humanities participated in the study and their reflections on their teaching experiences were analysed using NVivo 11. Solutions on meeting the challenges, good teaching practice and unresolved issues are discussed.

Keywords: higher education, inclusive teaching, International academics, new recruits, pedagogy

Introduction

"Hong Kong as an education hub" was cited in the University Grants Committee report (2010), and a forecast that "Asia will constitute about 70% of the global demand for higher education" in 2025 was also in the report (p. 55). There has been an ongoing strategy of recruiting international and mainland staff while promoting a good mix of international and local students in various programs in universities in the past two decades in Hong Kong (Lo & Ng, 2013; Lee, 2014). At the same time, internationalization means learning English (Choi, 2010), because many university students have had Chinese as the medium of instruction in school as part of the academic adaptation (Yu & Wright, 2017). The new four-year curriculum emphasizes an outcome-based education, and this has elicited many changes to the curriculum in both content and structure by local universities, including strategies on broadening science and arts education, bridging co-curricular activities, and outcome-based assessment to match international standards (Lanford, 2016). Despite the American style of liberation arts education being adopted at undergraduate level, many international staff have been recruited with new courses that were designed during the educational reform. This paper presents a study of new recruits to a local university in Hong Kong.

[*]Professional Consultant, the Chinese University of Hong Kong, Hong Kong.

New recruits may mean new to university teaching, new to teaching in Hong Kong but with a few years of teaching in university in other part of the world, or new to teaching in this university but have experience of teaching in Hong Kong. To acquire proficiency in university teaching, new recruits may start from their own learning experience as students. However, to educate 21st-century learners, they need to create a learning environment that focuses on student thinking and learning because they are living in an ever-changing world (Whitcomb, Borko, & Liston, 2009). In preparing for new recruits in the changing educational landscape, the central unit that provides the orientation and emerging pedagogies play a significant role in building both staff capacity to adapt in their transition to working in a new environment and have opportunities to reflect on effective pedagogies that invite inclusivity in the context of Hong Kong. This paper will discuss the outcomes of the dialogical approach with new recruits on formulating effective teaching practices in a university in Hong Kong.

New International Academics

The global higher education ranking system has a major influence on the promotion of internationalization in higher education, including mobility of students and teachers (Knight, 2013). This is because international staff can raise the international publication portfolios (Maringe & Sing, 2014). Educators have had a lot more opportunities to teach outside their home countries because globalization of higher education promotes international recruitment around the world, and many have experienced an initial adjustment phase (Jackson, 2008). There is a good assumption that new recruits know what academic life is and how to make a commitment to teaching and research and not anticipating any cultural shock (Foote, Wei, Monk, & Theobold, 2008). After arrival, new recruits will experience a period of re-establishment of a new identity through in-department and central support while experiencing the professional adjustments (Green & Myatt, 2011). This includes ways of living in a new country and various departmental and institutional practices, and some recruits may experience a steep learning curve even though academic sharing is a common goal when pursuing teaching and research excellence (Thomas & Malau-Aduli, 2013). Moving into a totally new teaching environment, new recruits need to adapt to differences in academic standards, grading systems and expectations of students on student–teacher relationships and their behavior in class (Phillips, Harris, Larson, & Higgins, 2009). They may be assigned to teach new subjects, to conduct mass lectures with a teaching lectern and be familiar with learning management systems. Nevertheless, internationalization is "the process of integrating an international, intercultural or global dimension into the purpose, functions or delivery of postsecondary education" (Knight, 2015, p. 3). Internationalization of the curriculum has created many opportunities for redesign and infusion of the new dimension of interdisciplinary, inclusive nature of multiculturalism. It is not merely what content is renewed, but how it is taught (De Vita & Case, 2003; Robson & Turner, 2007). This brings forth a test on a teacher's ability to teach in a

different culture (Hathaway, 2018) and deploy a variety of learning and assessment tasks that can embrace inclusivity (Maringe & Sing, 2014). Unsurprisingly, these teachers experience a heavy workload in which they have to make extra efforts in preparation but with very limited teaching experience (Green & Myatt, 2011).

Methodology

To cultivate a high-quality, trusting relationship, the author adopted a dialogic approach when having discussions with individual new recruits, because this could cultivate an openness as if having collegial communication through respectful listening and a sense of sharing (Gergen, McNamee, & Barrett, 2001; Shotter, 2006). The one-off, one-on-one engagement with new recruits offers collegial and intellectual discussion (Beech, Macintosh, & Maclean, 2010). An informal meeting was arranged with each participant, and there were some common questions on teaching and learning, including (1) when the staff member has started serving in the university, (2) the types of courses that the staff member taught, and (3) different ways of teaching with undergraduates and postgraduates that the staff adopted. While paying attention to their teaching approaches, the author made observations on expression of the challenges experienced. After the informal discussion, these staff would reflect on the discussion and make a reflection on their teaching and thoughts generated. Their reflections were used for analysis in this paper. Four themes emerged: opportunities, challenges, solutions and good practice after the reflections were coded with NVivo11.

Twenty-five new recruits from social sciences and arts and humanities were invited to participate in the study. There were five from Hong Kong (20%), eight from China (32%), eight from Asia (three from Korea, three from Taiwan, one from Japan, one from Vietnam; 32%), three from Europe (one from the United Kingdom, one from Spain, one from Sweden; 12%) and one from the United States of America (4%) (Table 1). However, many of them obtained their highest qualifications overseas as shown in Table 2: six graduated in Hong Kong (24%), five in Europe (three from the United Kingdom, one from Spain, one from Sweden; 20%) and 14 in the United States of America (56%). This means that 76% of new recruits have had international research and education experiences. Twenty-two of them were hired in the research track and three in the teaching track. Despite their employment records, all of them had teaching responsibilities. Names of the participants were replaced with pseudonyms in the *Findings and Discussion* section.

Table 1. Nationality of New Recruits
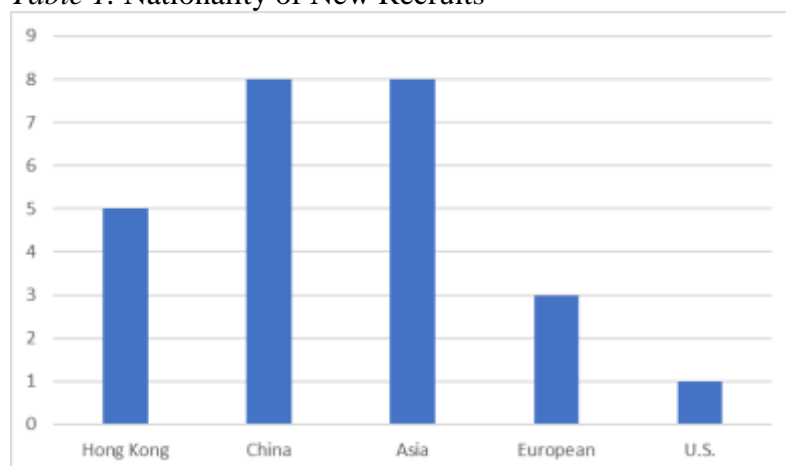
Table 2. Highest Qualification of New Recruits
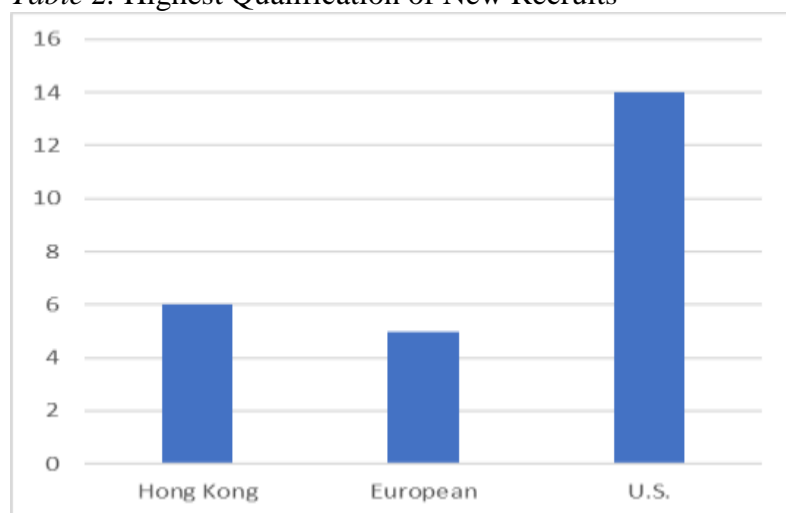
Findings and Discussion

Of the 25 new recruits, the majority obtained their highest research qualification from the United States of America. Thirteen were new to university teaching; nine had overseas teaching experience but would be unfamiliar with Chinese culture. Nevertheless, three of them had taught in other universities in Hong Kong and were very familiar with the learning behaviour of students in Hong Kong. In the reflection, there were many indicators that diversity of educational background, teaching and research experiences of these new recruits brought a rich teaching environment to the university. In the following, enriched learning opportunities, good practice, challenges and solutions were discussed.

Enriched Learning Opportunities

New international recruits have brought many new opportunities to local

students because of their personal overseas education and research experiences. More importantly, future graduates will be well prepared for their international graduate destination, given the variety of international teaching strategies experienced in class (Stone, 2006).

- "I start the first class with small-group and/or whole-class activities (e.g., sharing students' (and my) background and culture) to help students get acquainted, and build a sense of community with trust" [T11, social work].
- "I personally like a symposium-like teaching style: I prefer to provide guidelines to students; inspire them to brainstorm and discuss, before I step in to give them more concrete suggestions as to how to put the ideas into action (for research or application)" [T12, sociology].
- "I divide students into peer groups hoping they can help each other by reading each other's work from the first to final stage" [T14, gender studies].
- "Integrating photography with social media in student assignment…a creative process focused on the ability to generate ideas that are both innovative and functional…'rooftop debate' became a pleasant habit in my class" [T25, architecture].

Internationalization of the curriculum means that there is an international dimension to degree programs. More importantly, there is a need to help students to gain a global perspective and open-mindedness as global citizens. This means that they are exposed to an international focus and have had opportunities to develop intercultural competence through interaction with international staff and students on campus (Soria & Troisi, 2014) and the possibility of attending international events.

- "Having social work education and practice experiences in multiple countries/areas, I often use cases in different cultures and backgrounds as examples to enrich students' knowledge of human diversity" [T9, social work].
- "The topics range from glass ceiling in North America, foreign brides in Taiwan, domestic workers in Hong Kong, reproduction and parenthood in North Europe, manga and anime in Japan, beauty economy and sex work in China, and tied migration in US" [T17, gender studies].
- "A critical principle of my teaching was cultivating in students a global perspective and provide them an opportunity to reconsider their ingrained beliefs and perspectives…mostly used U.S. or Korean cases as example…I assigned a number of readings written by Western scholars" [T2, journalism and communication].
- "In the international monetary economics course, I take time to discuss the fragility of the Bretton Woods system, as well as the parallel between the 2001 Argentine crisis and the 2010 Greek crisis" [T3, economics].
- "An international collaborative event 'Learning from Barcelona' having field studies with a design exercise has become very popular. Students

discover the city of Barcelona through thematic field trips. These 'itinerant lessons' are guided by prominent Barcelona historians, designers and planners" [T25, architecture].

New recruits have created many enriched learning possibilities through a variety of teaching strategies and approaches while maximizing engagement with students intellectually, broadening their horizons and meeting their needs. However, this does not undermine them when facing challenges. To those who have had experience of teaching in Hong Kong, they still need to adapt to a new teaching environment in a different university. For those new to teaching and with non-Asian backgrounds, they may experience a steeper learning curve as they adjust to a new culture, systems and expectations while working in Hong Kong.

Challenges

The student population is predominantly local in Hong Kong universities. Table 3 shows the population of undergraduates, taught postgraduates and research postgraduates in 2017–18. The challenge for new international recruits is about student language proficiency, because many of them were educated in schools with Chinese as the medium of instruction, and they commonly experience difficulty in acquiring proficiency in English to comprehend disciplinary-based learning, producing academic writing with appropriate styles and standards (Evans & Morrison, 2011; Maringe & Jenkins, 2015).

Table 3. Student Population 2017–18
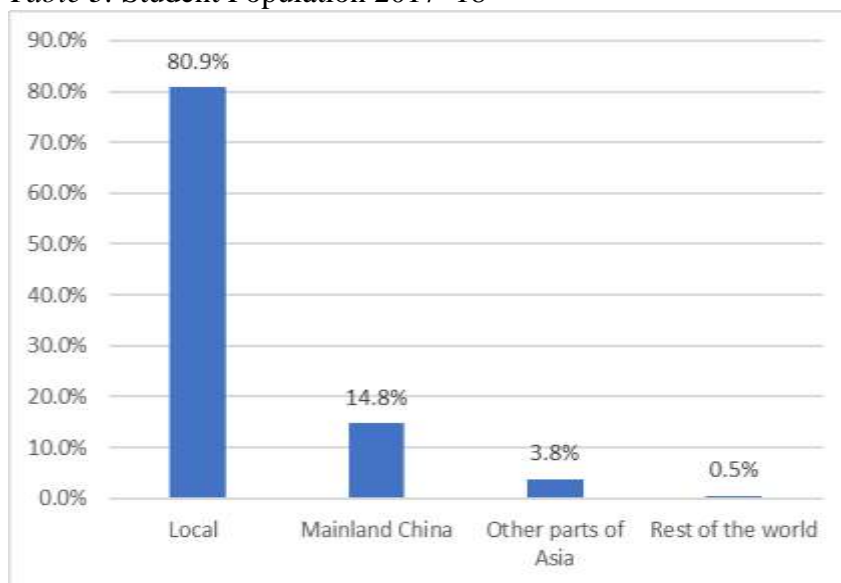
Diverse learning habits. Chinese learners have some common characteristics, such as being passive, and being reluctant to engage in critical thinking, argument and discussion. They regard teachers as authority figures but may be more dependent on them as mentors or pastoral counsellors who provide support and moral guidance (Gao & Watkins, 2002). Generally, they do not prefer participatory

modes of learning, particularly those less successful groups (Robson & Turner, 2007), and this may be due to a variety of reasons, including a lack of required language proficiency, motivation and avoidance of embarrassment (Cheng, 2000). Nevertheless, undergraduates show a tendency to depend on educators (Gan, 2009).

Around 76% of new recruits have had education and research experience abroad but have very limited knowledge of the student learning culture in Hong Kong. Those who have taught in the USA note that Hong Kong students seem relatively passive, are good at copying teacher's work to their notebooks, and are reluctant to share their ideas or opinions in class. However, some students would prefer to approach their teachers at break time or at the end of a class. Some students may have difficulty to connect theories and real-world practice, but they often highlight this in their PowerPoint slides in preparation for examination. Hong Kong students tend to finish group assignments before the deadline for submission but spend less preparation time compared with mainland China students. They may not express their satisfaction or dissatisfaction directly to their teachers, but they make comments on teaching evaluation at the end of all courses.

On the other hand, mainland China students come to study in Hong Kong because scholarships are available, the world ranking and reputation of universities in Hong Kong is high, and it is convenient to go back to China for Chinese festivals; and they perceive the importance of acquiring international qualifications (Li & Bray, 2007). Moreover, they expect educators to deliver content and transfer information in the way of teaching in China (Tavakol & Dennick, 2010). Nevertheless, they exercise self-managed individual learning after class as a habit in preparation for achieving a pass in the language proficiency test, which is nationwide and is compulsory (Gan, 2009). Mainland China students are perceived as non-local students even though they share the same cultural background; it is because the educational experiences of mainland China students are very different from those in Hong Kong. Many mainland students like studying by themselves. They often sit in the front of a class, prepare their assignments well before deadlines and pay high attention to detail in their work. Language proficiency varies among mainland China students, some of whom can speak as fluently as a native while others can barely pronounce English words correctly (Hodgson, Lam & Chau, 2011).

The learning habits of home students may be different from non-local students, and this can bring challenges to teachers to enable a congenial learning environment (Robson & Turner, 2007). This means that Hong Kong students and mainland China students do not normally group together in group discussions or in group projects and assignments. Teachers may consider ways to mix them collectively in class activities and assessment tasks.

Teaching undergraduates and postgraduates. New recruits may have to teach both undergraduates and postgraduates in the university. However, one teacher noted that "I unrealistically asked students to reflect on traditional research methods in social science from the angle of feminist epistemology that I taught in my postgraduate class" [T17, gender studies]. Undergraduates find the content too

advanced for them. Differentiation of learning challenges between undergraduates and postgraduates may seem obvious. However, research-tracked recruits comment that the gap in prior knowledge among learners in a master program is often large because those programs do not require students to have a prior undergraduate qualification. It is not uncommon to have a part-time master program with some students having years of professional working experience while others are new to the discipline but are aiming to acquire a master qualification.

Solutions

Addressing challenges in teaching in a university, educators need to focus on the needs of both local and non-local students. Local students expect teachers to be good moral models, warm-hearted and friendly, while non-local students expect their teachers to set interesting learning experiences (Walker & Dimmock, 2000). They need to have strategies to assist learners with different expectations.

Teaching and learning. Among an abundance of teaching methods, the use of real-world problems is one of the best methods to enable students to build critical-thinking skills while learning to solve problems (Halpern, 1999; Dwyer, Hogan, & Stewart, 2014). It is particularly important to expose students in undergraduate programs to such problems because they lack knowledge and hands-on skills at first, but they will need such skills in the real world. Moreover, educators play a central role not only in the cognitive domain but also in building a rapport with students and subsequently higher engagement with students, producing improved academic performance (Klem & Connell, 2004) whereas various student community activities can contribute to personal growth (Strayhorn, 2008).

- "Choose simple real-world examples" [T19, economics].
- "Connect concepts with real-life situations and real-life experience" [T15, nursing].
- "The design thinking process invites students to frame the puzzle or the problem as a mystery, involving the students more creatively in the problem-solving process" [T5, Centre for China Studies].
- "Enhancing students' motivation to learn through actively involving students in the learning process and creating a supportive learning environment" [T9, social work].
- "Education as a means to provide students with a basis of general knowledge and skills that enable independent critical thinking and promote personal self-growth…it is a continuous process that does not end with the mastery of a specific set of skills in the classroom or future workplace; it involves change not only in knowledge but also in personal beliefs and attitude" [T7, Independent Learning Centre].

Engaging Asian students. It is possible to engage Asian learners, but this means that educators need to adopt different strategies, including having students

provide anonymous responses (Liu, 2009), using digital devices for voluntary and involuntary activities (Baker Lusk, & Neuhauser, 2012), having more casual communication with them (Lee, 2009) and possibly assigning students roles and responsibilities (Mckay and Kember, 1997). Nevertheless, students tend to stay with peers from their own common cultural background despite the possibility of positive cross-cultural experiences happening (Montgomery, 2009). Therefore, educators can create opportunities for cross-cultural activities in the curriculum that focus on maximizing talents through inclusivity.

- "In the US, you can raise open questions and the students would not shy away from addressing them in public, while here at CUHK it would be better if you ask them to express their opinions or ideas anonymously" [T12, sociology].
- "I use 'hand-raising' function and 'pick or random' function with an interactive class system and this can enable more effective warming-up interactions" [T2, journalism and communication].
- "Informal conversation. I make efforts to be a person who comes to the classroom first and leaves the classroom last, because these efforts provide me with invaluable opportunities to have an informal conversation with students before and after classroom hours, which helps me to obtain genuine feedback from students" [T11, social work].
- "Students posting questions and peer group responses to one of the questions" [T1, social science].
- "Each group leads class discussion regarding a topic of the assigned week…tracking their participation, and reflecting it on their grade" [T2, journalism and communication].
- "There are a total of eight students in a group, three of whom are discussants to search for literature and books to provide information, two are respondents who criticize information, one student is secretary and is responsible for taking minutes, one student is a presenter, who will be reporting the results on behalf of the group, and one student is an observer, working with the secretary to prepare presentation slides" [T22, nursing].
- "Mixing students in groups [local/international] by alphabetical order in surnames" [T1, social science].
- "Each student to select three most-preferred topics out of six youth problems planned to be covered in the class, and then group students based on their preference…eases invisible tensions that sometimes arise between the two groups of students [Hong Kong vs mainland], suggesting that it is an effective strategy for promoting an 'inclusive' classroom" [T11, social work].

Differentiating undergraduate and postgraduate teaching. New recruits may be assigned to teach both undergraduate and postgraduate courses, and the less experienced teachers may be set tasks that may be perceived as difficult for undergraduates because the foundational knowledge has just been established. On the other hand, experienced educators will create a collegial learning environment

such that experienced adult learners can provide support to those with limited knowledge or skills (Hodgson & Chau, 2015).

- "Undergraduate course, I have to present concepts and theories step by step in a very logical way, pay more attention to their application to practical aspects. Graduate-level course, I tend to raise mismatches between theories and facts to motivate the students' discussion" [T19, economics].
- "I have proactive public engagement with partners from international organizations, business sectors, and national think tanks. These networks and experiences allow me to show my undergraduate students how to apply theories taught in class in real-time situations whereas I would connect postgraduate students to related real-world projects" [T20, global studies].
- "Postgraduate students are mature and experienced. I encourage students to share their difficult encounters with clients in their work setting. Through their sharing, students without social work experience can obtain a sense of the practicalities of social work practice" [T13, social work].

Good Teaching Practice

Experienced faculty members, though new to the university, have already adopted sound pedagogies in their disciplines, while some develop engaging strategies through reflection on actions (Mälkki & Lindblom-Ylänne, 2012). Among the new recruits, some have had overseas teaching experience, and some reflect on formal and informal student feedback, both positive and negative. Some good teaching practices have emerged among these new recruits:

1. Attend to short attention span:
   "Adopt a variation of instruction modality" [T8, social science].
2. Enable development of transferable skills:
   "Cultivate students' intellectual capability, have them appreciate the peer-review exercise, such that they gain confidence in expressing opinions" [T14, gender studies].
3. Engage student interest in assessment tasks:
   - "I designed some real-world projects, such as how to make a Facebook fan page more popular by using the data students have from Facebook instead of teaching them how to code and how to analyse the data using statistical models" [T23, journalism and communication].
   - "Having role play of real estate developers in the public auction process, students are asked to provide a price that is acceptable to them for relocation. The 'auction' function enables them to provide their acceptable price levels without being influenced by their 'neighbours', as the rule of the game is to accept relocation and game over if the majority of residents agree the price" [T24, geography and resource management].
4. Link research and teaching:

- "I bring my academic research into class….I also strive to incorporate my research on economic disparities and welfare policies into class, which helps students to become aware of the fact that social workers engage in a broad range of activities within multiple types of settings in order to promote social justice" [T11, social work].
- "In the theory-based courses, students are required to complete an academic research paper. The process took them from constructing a theoretical framework to designing a research method, from data collection/analyses to data interpretation, and develop a term paper that is now under review" [T2, journalism and communication].

5. Provide prompt feedback to students:
   - "Students can come to seek my comments on the final draft of term papers a few days before deadline without being graded. It encourages those students who want to perfect their work by having a final touch-up in light of my feedback as a series of pedagogical dialogues between students and I in order to cater for learning diversity, where guidance and improvement become possible" [T10, gender studies].
6. Pay attention to student feedback:
   - "Students welcome more multimedia learning materials and want to have more discussion time between them, particularly for a course with theories that are too abstract to comprehend. I adjust more time for discussion, and provide more multimedia resources" [T4, economics].
   - "I replace quizzes with more flexible discussion and writing tools. These tools are designed to encourage careful deliberation, effective communication, and persuasive presentation, with skills of initiating dialogue, conducting debates, and directing discussion flows with respect to others" [T17, gender studies].

Unresolved Challenges

The biggest drawbacks for non-local new recruits is working in an environment with a language unknown to most of them. This makes it neither possible to access to know resources in the local context, which is traditional Chinese, nor is it possible to communicate freely with local students and the public through the local language. Although they can use materials in English and draw in international examples in their teaching, some local undergraduate students, particularly those less articulate, may not take the opportunity to respond in class or raise questions without concern of embarrassment. Although students may be exposed to international perspectives, they may not be able to act local and think global because there is a shortfall of local connections between different sectors in Hong Kong. Thus, despite new recruits having had an international profile and international collaborators, it takes time to establish a local reputation and connections (Green & Myatt, 2011).

In addition, the university has had a bilingual policy of university education, and it is not a surprise that some students express the need to use their first

language to learn and to practise professional knowledge in their teaching evaluations, despite these courses being set as having English as the medium of instruction. Course-end teaching evaluations are marked down because students prefer to be able to learn through their native language (Bodycott & Walker, 2000).

Conclusion

Universities will benefit from new international and local recruits because they bring forth new and diverse approaches to teaching in the classroom while maintaining critical reflection and practising effective teaching. This can be a win–win strategy for institutions to learn from international academics about ways to improve existing pedagogy and curriculum practices (Hsieh, 2012). These staff can be a special resource in universities because they provide extensively rich learning opportunities for home learners such that they can build global perspectives and be better prepared for their future careers (Standley, 2015). The majority of students may not have the opportunity to study abroad. Therefore, it is essential for them to be well prepared for intercultural competence and global perspectives through the curriculum so that they can professionally and personally fit into a global economy and be competent locally (Doyle et al., 2010). For local recruits, the extensive teaching and research experience can create abundant opportunities for students to be connected with communities for authentic learning experiences.

Nevertheless, new international recruits may experience a lot of stress with expectations on excellence in teaching and research (Tynan & Garbett, 2007). As they adjust to the local culture and practices, they may experience dissonance if anticipation is mismatched with reality (Hathaway, 2018). They need to see the complexity of local undergraduate students with English as their second language to adopting academic proficiency in written and oral communications (Kennedy, 2002). As there are many new approaches to teaching Hong Kong students and mainland China students, understanding their differences is necessary. Internationalization at home can connect not only global scholars but also opportunities to have creative and effective pedagogies that can instill student curiosity through their passion to pursue excellence in educating future graduates.

References

Baker, W. M., Lusk, E. J., & Neuhauser, K. L. (2012). On the use of cell phones and other electronic devices in the classroom: evidence from a survey of faculty and students. *Journal of Education for Business*, *87*(5), 275–289. https://doi.org/10.1080/08832 323.2011.622814

Beech, N., Macintosh, R., & Maclean, D. (2010). Dialogues between academics and practitioners: the role of generative dialogic encounters. *Organization Studies*, *31*(9–10), 1341–1367. https://doi.org/10.1177/0170840610374396

Bodycott, P., & Walker, A. (2000). Teaching abroad: lessons learned about inter-cultural understanding for teachers in higher education. *Teaching in Higher Education*, *5*(1), 79–94. https://doi.org/10.1080/135625100114975

Cheng, X. (2000). Asian students' reticence revisited. *System*, *28*(3), 435–46. https://doi.org/10.1016/S0346-251X(00)00015-4.

Choi, P. K. (2010). "Weep for Chinese university": A case study of English hegemony and academic capitalism in higher education in Hong Kong. *Journal of Education Policy*, *25*(2), 233–252. https://doi.org/10.1080/02680930903443886.

De Vita, G., & Case, P. (2003). Rethinking the internationalisation agenda in UK higher education. *Journal of Further and Higher Education*, *27*(4), 383–398. https://doi.org/10.1080/0309877032000128082.

Doyle, S., Gendall, P., Meyer, L., Hoek, J., Tait, C., Mckenzie, L., & Loorparg, A. (2010). An investigation of factors associated with student participation in study abroad. *Journal of Studies in International Education*, *14*(5), 471–490. https://doi.org/10.1177/1028315309336032

Dwyer, C.P., Hogan, M.J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity*, *12*, 43–52. https://doi.org/10.1016/j.tsc.2013.12.004

Evans, S., & Morrison, B. (2011). Meeting the challenges of English-medium higher education: the first-year experience in Hong Kong. *English for Specific Purposes*, *30*(3), 198–208. https://doi.org/10.1016/j.esp.2011.01.001.

Foote, K.E., Wei, L., Monk, J., & Theobold, R. (2008). Foreign-born scholars in US universities: issues, concerns and strategies. *Journal of Geography in Higher Education*, *32*(2), 167–178. https://doi.org/10.1080/03098260701731322

Gao, L., & Watkins, D. A. (2002). Conceptions of teaching held by school science teachers in P.R. China: identification and cross-cultural comparisons. *International Journal of Science Education*, *24*(1), 61–79. https://doi.org/10.1080/09500690110066926

Gan, Z. (2009). "Asian learners" re-examined: an empirical study of language learning attitudes, strategies and motivation among mainland Chinese and Hong Kong students. *Journal of Multilingual and Multicultural Development*, *30*(1), 41–58. https://doi.org/10.1080/01434630802307890.

Green, W., & Myatt, P. (2011). Telling tales: a narrative research study of the experiences of new international academic staff at an Australian university. *International Journal for Academic Development*, *16*(1), 33–44. https://doi.org/10.1080/1360144X.2011.546219.

Gergen, K.J., McNamee, S., & Barrett, F. (2001). Toward transformative dialogue. *International Journal of Public Administration*, *24*, 697–707. https://doi-org.easyaccess1.lib.cuhk.edu.hk/10.1081/PAD-100104770.

Halpern, D.F. (1999). Teaching for critical thinking: helping college students develop the skills and dispositions of a critical thinker. *New Directions for Teaching and Learning*, *80*, 69–74. https://doi.org/10.1002/tl.8005.

Hathaway, T. (2018). Negotiating transitions in academic identity: teacher or researcher? In A. Hosein, N. Rao, C.S.-H. Yeh, I.M. Kinchin (eds), *Academics' International Teaching Journeys: Personal Narratives of Transitions in Higher Education*, pp. 93–108. Bloomsbury Publishing. ISBN: HB: 978-1-4742-8977-1.

Hodgson, P., & Chau, J. (2015). Optimizing interactions to enhance English competence for university students as global learners: A tale of two universities in Hong Kong. *International HETL Review*, *5*(7). https://bit.ly/2SAKriW.

Hodgson, P., Lam, P., & Chow, C. (2011). Assessment experience of first-year university students: dealing with the unfamiliar. *Proceedings of Enhancing Learning*

Experiences in Higher Education: International Conference (digital). https://www.cetl.hku.hk/conference2010/pdf/Hodgson.pdf.

Hsieh, H. (2012). Challenges facing Chinese academic staff in a UK university in terms of language, relationships and culture. *Teaching in Higher Education*, *17*(4), 371–383. http://dx.doi.org/10.1080/13562517.2011.641001.

Jackson, J. (2008). Globalization, internationalization, and short-term stays abroad. *International Journal of Intercultural Relations*, *32*(4), 349–358. https://doi.org/10.1016/j.ijintrel.2008.04.004

Kennedy, P. (2002). Learning cultures and learning styles: myth-understandings about adult (Hong Kong) Chinese learners. *International Journal of Lifelong Education*, *21*(5), 430–445. https://doi.org/10.1080/02601370210156745.

Klem, A.M., & Connell, J.P. (2004). Relationships matter: linking teacher support to student engagement and achievement. *Journal of School Health*, *74*(7), 262–273. https://doi.org/10.1111/j.1746-1561.2004.tb08283.x.

Knight, J. (2013). The changing landscape of higher education internationalisation – for better or worse? *Perspectives: Policy and Practice in Higher Education*, *17*(3), 84–90. https://doi.org/10.1080/13603108.2012.753957.

Knight, J. (2015). Updated definition of internationalization. *International Higher Education*, *33*, 2–3. https://doi.org/10.6017/ihe.2003.33.7391.

Lanford, M. (2016). Perceptions of higher education reform in Hong Kong: a glocalization perspective. *International Journal of Comparative Education and Development*, *18*(3), 184–204. http://dx.doi.org/10.1108/IJCED-04-2016-0007.

Lee, G. (2009). Speaking up: six Korean students' oral participation in class discussions in US graduate seminars. *English for Specific Purposes*, *28*(3), 142–156. https://doi.org/10.1016/j.esp.2009.01.007.

Lee, H.M. (2014). Hong Kong higher education in the 21st century. *Hong Kong Teachers' Centre Journal*, *13*, 15–33.

Li, M., & Bray, M. (2007). Cross-border flows of students for higher education: push–pull factors and motivations of mainland Chinese students in Hong Kong and Macau. *Higher Education*, *53*(6), 791–818. https://doi.org/10.1007/s10734-005-5423-3.

Liu, J. (2009). From learner passive to learner active? The case of Chinese postgraduate students studying marketing in the UK. *International Journal of Management Education*, *7*(2), 33–40.

Lo, W. Y.W., & Ng, F.S.K. (2013). A critical reflection on internationalization of higher education in Hong Kong: the search for a cosmopolitan alternative. *Asia Pacific Journal of Educational Development* (APJED), *2*(1), 37–46. http://dx.doi.org/10.6228%2fAPJED.2013.0201.04.

Mälkki, K., & Lindblom-Ylänne, S. (2012). From reflection to action? Barriers and bridges between higher education teachers' thoughts and actions. *Studies in Higher Education*, *37*(1), 33–50. https://doi.org/10.1080/03075079.2010.492500.

Maringe, F., & Sing, N. (2014). Teaching large classes in an increasingly internationalising higher education environment: pedagogical, quality and equity issues. *Higher Education*, *67*(6), 761–782. https://doi.org/10.1007/s10734-013-9710-0.

Maringe, F., & Jenkins, J. (2015). Stigma, tensions, and apprehension. *International Journal of Educational Management*, *29*(5), 609–626. https://doi.org/10.1108/IJEM-04-2014-0049.

Mckay, J., & Kember, D. (1997). Spoon feeding leads to regurgitation: a better diet can result in more digestible learning outcomes. *Higher Education Research & Development*, *16*(1), 55–67. https://doi.org/10.1080/0729436970160105.

Montgomery, C. (2009). A decade of internationalisation: has it influenced students' views of cross-cultural group work at university? *Journal of Studies in International Education*, *13*(2), 256–270. https://doi.org/10.1177/1028315308329790.

Phillips, D., Harris, G., Larson, M., & Higgins, K. (2009). Trying on-being in-becoming. Four women's journey(s) in feminist poststructural theory. *Qualitative Inquiry*, *15*(9), 1455–1479. http://dx.doi.org/10.1177/1077800409347097.

Robson, S., & Turner, Y. (2007). "Teaching is a co-learning experience": academics reflecting on learning and teaching in an "internationalized" faculty. *Teaching in Higher Education*, *12*(1), 41–54. https://doi.org/10.1080/13562510601102115.

Shotter, J. (2006). Understanding process from within: an argument for "withness"-thinking. *Organization Studies*, *27*(4), 585–604. https://doi.org/10.1177/0170840606 062105.

Soria, K.M., & Troisi, J. (2014). Internationalization at home alternatives to study abroad: implications for students' development of global, international, and intercultural competencies. *Journal of Studies in International Education*, *18*(3), 261–280. https://doi.org/10.1177/1028315313496572

Standley, H.J. (2015). International mobility placements enable students and staff in higher education to enhance transversal and employability-related skills. *FEMS Microbiology Letters*, *362*(19), 1–5. https://doi.org/10.1093/femsle/fnv157.

Stone, N. (2006). Internationalising the student learning experience: possible indicators. *Journal of Studies in International Education*, *10*, 409–413. https://doi.org/10.1177/1028315306287633

Strayhorn, T. L. (2008). How college students' engagement affects personal and social learning outcomes. *Journal of College and Character*, *10*(2), 1–16. doi: https://doi.org/10.2202/1940-1639.1071.

Thomas, S.L., & Malau-Aduli, B.S. (2013). New international academics' narratives of cross-cultural transition. *International Journal of Higher Education*, *2*(2), 35–52. https://doi.org/10.5430/ijhe.v2n2p35.

Tavakol, M., & Dennick, R. (2010). Are Asian international medical students just rote learners? *Advances in Health Sciences Education*, *15*(3), 369–77. https://doi.org/10.1007/s10459-009-9203-1.

Tynan, B., & Garbett, D. (2007). Negotiating the university research culture: collaborative voices of new academics. *Higher Education Research & Development*, *26*(4), 411–424. https://doi.org/10.1080/07294360701658617.

Walker, A., & Dimmock, C. (2000). Mapping the way ahead: leading educational leadership into the globalised world, *School Leadership & Management*, *20*(2), 227–233. https://doi.org/10.1080/13632430050011443.

Whitcomb, J., Borko, H., & Liston, D. (2009). Growing talent: promising professional development models and practices. *Journal of Teacher Education*, *60*(3), 207–212. https://doi.org/10.1177/0022487109337280.

University Grants Committee (2010). *Aspirations for the Higher Education System in Hong Kong: Report of the University Grants Committee*. Hong Kong: University Grants Committee.

Yu, B., & Wright, E. (2017). Academic adaptation amid internationalisation: the challenges for local, mainland Chinese, and international students at Hong Kong's universities. *Tertiary Education and Management*, *23*(4), 347–60. https://doi.org/10.1080/13583883.2017.1356365.

The Relationship between Teachers' Perceptions of Emotional Labor and Teacher Burnout and Teachers' Educational Level

By Racheli Zaretsky[*] & Yaacov J Katz[†]

The purpose of this research was to examine the relationship between teachers' perceptions of emotional labor, teacher burnout and teachers' educational levels. The research sample consisted of 170 female Haredi (religiously ultra-orthodox and observant) teachers working in schools throughout Israel. The data were collected using the Emotional Labor of Teaching Scale (TELTS) and Teacher Burnout Questionnaire, and a demographic questionnaire formulated for the present study. The data collected in the study were processed using SPSS software. The two research hypotheses, namely a) teachers characterized by lower levels of burnout will use the natural acting technique associated with emotional labor whereas teachers typified by higher level of burnout will utilize the surface acting or deep acting emotional labor techniques; and b) teachers with higher levels of education will tend to use the natural acting technique associated with emotional labor in contrast to teachers with lower levels of education who will tend to utilize the surface acting or deep acting emotional labor techniques, were analyzed by ANOVA and MANOVA procedures. The main findings indicated that both teachers' burnout levels and teachers' levels of education differentially affected the use of the specific emotional labor techniques (surface, deep or natural acting). The conclusion of this study is that various variables can influence teachers' use of techniques associated with emotional labor. Changes in such variables, such as teachers' level of burnout and level of education, may improve management of emotions in teaching and improve their well-being.

Keywords: emotional labor, level of burnout, level of education.

Introduction

"Teaching Is Emotional Labor" (Larrivee, 2012, p. 37). *"Emotions are dynamic parts of ourselves, and whether they are positive or negative, all organizations, including schools, are full of them"* (Hargreaves, 1998, p, 835). Thus, emotional labor is perceived as a central and major factor associated with the wellbeing of members of the workforce in general and particularly those in the teaching profession as is evident in the above quotations. This paper will address the relationship between emotional labor of teachers and their levels of burnout as well as with their levels of education.

[*] PhD Student, School of Education, Bar-Ilan University, Israel.
[†] Professor Emeritus, School of Education, Bar-Ilan University, Israel.

Emotional Labor

Feeling Rules

According to the Meriam-Webster dictionary (2017), emotion is: "*A conscious mental reaction (as anger or fear) subjectively experienced as a strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body"*. The expression of feelings has many diverse social functions: it is related to gathering of information, influences social interactions and shapes interpersonal relationships with others (Wong et al., 2013). This relationship, between emotion and the social context, creates *feeling rules* (Hochschild, 1983; Hochschild, 2012) which can be described as norms that relate to the way a culture perceives the suitability or appropriateness of ways to express emotions in specific social interactions, including the degree and intensity of which expressions are expressed (Ekman & Friesen, 1975; Matsumoto & Hwang, 2013). Feeling rules differ from society to society and from culture to culture (Matsumoto & Hwang, 2013). They change according to the specific context (Moran, Diefendorff, & Greguras, 2013) and in accordance with the changes or transformations that take place (Brennan, 2006). Moreover, the reminders and the sanctions change from group to group and from situation to situation (Hochschild, 1983; 2012). Nevertheless, there are those who assert that even in the same given social context, and at the same point in time, feeling rules are dependent upon the expression of feelings or emotions, in terms of the subjective interpretation that everyone perceives in a specific situation (Ginat, 2011).

Emotional Regulation

Feeling rules require the individual, in many daily situations of his or her life, to try to hide certain emotions or feelings that are not congruent with his or her true or real feelings. This is done to minimize the discrepancy or gap that exists between the authentic feeling (What I am feeling) and the ideal feeling (What I am supposed to feel) as described by Hochschild (1983; 2012). This process of presentation or repression of emotion, via facial expressions, bodily gestures and or verbal expressions, which are fashioned to match the situation and the context, is termed "emotional regulation" in the research literature (see Hochschild, 1983; Gross, 1998; Grandey, 2000).

It is possible to identify two kinds of strategies for emotional regulation: a) an adaptive strategy, which includes re-examination or evaluation of the situation, problem solving or acceptance of the emotions, and b) a non-adaptive strategy, which is expressed through conscious repression of the emotion due to the desire to minimize negative tensions. It has been found that the use of adaptive strategies is related to lower levels of anxiety and depression, while the use of non-adaptive strategies care related to higher levels of these phenomena (Subic-Wrana et al., 2014). Different situations dictate the use of differential

strategies for the regulation of emotion, since the rules for emotional expression change from context to context.

Emotional Labor

In the context of organizational and business employment, emotional regulation is defined as emotional labor. This concept was first introduced in 1983 by the sociologist, Arlie Hochschild (1983) in her groundbreaking volume, and was defined by her as emotional regulation that functions to the norms of the area of business/profession/vocation or of the organization. In this volume the consequences of emotional regulation in the work place in the service sector were examined vis-à-vis the organizational level as well as on the psychological level of the individual worker (Grandey, 2015).

Techniques Associated with Emotional Labor

Emotional labor is dependent upon active techniques for change or for the creation and matching of the feelings or emotions that are expressed in the context of paid work. Hochschild (1983; 2012) identified two central techniques of emotional labor through which workers in fields that combine work with customers accept the rules that relate to the demonstration of emotions that are dictated to them by the organization employing them: (a) "surface acting" – the expression of an ideal emotion that is not in line with the authentic emotion and (b) "deep acting" – transformation of the ideal emotion into an authentic emotion with the goal of "toeing the line" concerning expression of emotions that is demanded by the organization. After a decade, a third technique was identified by Ashforth & Humphrey (1993), "natural acting". This is the expression of genuine emotion. Here the individual spontaneously demonstrates authentic emotion which matches or suits the ideal emotion, without the need to perform any manipulation on the emotions.

Emotional Labor in Teaching

Emotional Labor in Teaching: State of the research

Teaching serves as one of the many professions that has been researched in the context of emotional labor. In fact, it is one of the most studied (Tsang, 2014). Over the years, many studies have examined a variety of aspects connected to emotional labor in teaching, including many recent studies (see Hagenauer & Volet, 2014; Levine Brown et al., 2014; Ching-Sheue, 2015; Sahin, 2015) and meta-analyses (for example, Tsang, 2014; Ye & Chen, 2015). This is in light of the strengthening, over the last few decades, of the perception that the profession of teaching belongs to the service sector, which perceives the system of relations between teachers and students as relations between

service-providers and customers (Oplatka, 2011; Iltaf & Gulzar, 2013) and views the teacher as a figure whose role is to create an emotional change in the attitudes of students toward studying, learning and education (Basim, Begenirbaş, & Yalçin, 2013).

The study of emotional labor among teachers is found throughout the world and includes studies that were undertaken in many and different countries. Pfister (2015), who surveyed the academic literature that deals with emotional labor in the context of schools, found that these studies concentrated on countries in North America, Europe, Asia and Australia, while she did not find research that was undertaken in Africa or South America. She notes that only one study was undertaken in Israel in 2007 (see Pfister, 2015, pp. 32-35 for a table that presents the studies according to country).

The study of emotional labor among teachers and the research on its impact and consequences on their psychological and emotional welfare and well-being is one of the topics that has been well-researched in the field of emotional labor (Tsang, 2014). There are several reasons for this, which include: (1) emotional labor influences the efficiency and efficacy of the pedagogical-educational work of the teachers (Roberts, 2011; Sutton, 2004); (2) emotional labor influences the self-image of the teachers since, for them, the classroom becomes their main source of self-esteem, satisfaction and vulnerability (Roberts, 2011). Nevertheless, it has been found that the study of emotional labor in the teaching profession has a unique character, which differentiates it from the other service professions.

According to Larrivee (2012, p. 37) "teaching is emotional labor". To begin with, interaction with students is usually much longer, continuous and more powerful than interaction with clients in different service industries. Secondly, the frequency and degree of problematic behaviors (beginning with chatting during lessons and not paying attention through rude comments, verbal and physical violence and absences) to which teachers are exposed in schools are significantly higher than in other professions. This does not only disrupt the process of teaching, it also appears to be one of the central factors contributing to emotional exhaustion. This is because it is difficult to for teachers control feelings of frustration, anger and other negative emotional expressions when students continually to ignore their instructions or behave in ways that disrupt the class (Wróbel, 2013).

Hagenauer & Volet (2014) point to two main reasons for the high interest in studying emotional labor in the field of teaching. (1) It was found that teachers experience a wide range of emotions, including positive emotions (such as pleasure, pride and satisfaction) and negative ones (such as anxiety, frustration, anger, sadness and embarrassment) and that these emotional experiences influence the well-being of the teachers, their satisfaction from the profession of teaching and the degree of danger or risk of them becoming burnt-out and leaving the profession. (2) It was found that the emotions of the teachers or the character of their emotional relationships with their students have an influence of the character of their teaching. As a result, they impact the quality of the product of teaching.

However, as opposed to business organizations, the consequences of emotional labor in teaching are not limited to the organization or solely to the worker: there are far-reaching consequences for society, that impact different circles including the community, the country and even the world and, so it seems, the future generation as well. As a result, the study of emotional labor in the field of teaching in schools is not only theoretical. Its great importance is derived from the need to prevent negative influences on the educational system and on entire society (Sun, 2013). Moreover, the reality in Israel, like in other countries, positions the teachers as service providers, the students (and the parents) as clients or customers, and the school administration, the supervisor, the school board and the Ministry of Education as directors and employers (Oplatka, 2009; Ginat, 2011).

This phenomenon is even more conspicuous due to the policy of opening areas of registration of schools and the creation of inter-school competitions in cities. As a result, new feeling rules have been created that are expressly designed to keep students in school. According to these rules, it is expected that the teacher, among others, will avoid expressing emotions that are judged to be overly negative toward students, for fear that the latter will leave the school (Oplatka, 2011). As a result, we can summarize that the management of emotions is an inseparable part of role requirements in the teaching profession (Levine Brown, 2011).

Emotional Labor Techniques used by Teachers

As mentioned above, the emotional labor of teachers deals with the emotions that arise from their professional work with their students, their superiors, their head-teachers, their supervisors and Ministry of Education officials. The current research focuses on one front of the emotional labor of the teacher, that related to their relationship with their students. Results of research studies have indicated the existence of unique behaviors in the field of teaching, in the context of each one of the techniques of emotional labor. For example, Levine Brown (2011) has detailed the following behaviors in a comprehensive research study:

a. Surface acting: the teacher makes an effort to conceal from the students emotions that are considered to be undesirable: "In my work with students, I act differently from the way that I actually feel"; "Even when I am angry or worried, I cause people in my environment to think that my mood is good"; "In order to do my job, I pretend that I actually feel the emotions that people expect me to express/demonstrate".

b. Deep acting: the teacher makes an effort to authentically feel the emotions expected from himself or herself and to demonstrate them to the students: "I make an effort to really feel the emotions that I am expected to demonstrate at work"; "I really try to feel the emotions that they expect from me"; "I work hard in order to feel the emotions that are expected from me to demonstrate at work".

    c. Natural acting: the teacher manages her or his emotions in accordance with her or his authentic feelings, and this is what she or he demonstrates to her or his students in the most natural manner, without the need "to put on a show". "The emotions that I show my students match my real feelings"; "The emotions that I show my students come to me naturally".

Burnout

Burnout: Definition and State of the Research

Burnout is a term used in many fields in the social sciences and in administration sciences. Herbert Freudenberg (1974) during his work as a psychiatrist, examined the gradual process of "the emptying out of emotions" and the loss of motivation and obligation to the job, which is accompanied by a variety of psychological and physical symptoms which develop as a response to interpersonal stress factors at work. He defined burnout as a state of physical, emotional and psychological weariness or tiredness, with symptoms that include a decrease in the level of achievement, de-personalization and a decrease in the level of interest in work. Freudenberg represents the pioneers dealing with this concept, who began studying the phenomenon in the mid-1970s in the United States. The main contribution made by Freudenberg and his counterparts was that of naming, describing and presenting the insight about this phenomenon, which turned out to be quite common. His conclusions were based on the results of various qualitative studies that included personal interviews of workers in different vocations in which they provided services to individuals who needed them. These studies focused on the phenomenon of burnout among the service providers (Maslach, Schaufeli, & Leiter, 2001).

Burnout in Teaching

The teaching profession is thought of as being one of the professions most characterized by burnout. The research on this topic began in the early (Gavish & Friedman, 2010) and has continued to this day. Research findings have shown that the main reason for burnout of teachers is related to personal characteristics of teachers. It appears that there are several personality characteristics that can increase the danger or risk of burnout. Among these characteristics, one finds the tendency for neuroticism, such as anxiety, lack of security or nervousness, the tendency toward passivity, introversion, worry about what others are thinking, obsessive thoughts about mistakes that were made in one's teaching, self-criticism and a low threshold for frustration (Larrivee, 2012). Another personal characteristic that was found to predict the risk of burnout among teachers is little experience in the field of teaching (Larrivee, 2012).

There are far-reaching consequences of burnout on teachers, from a decrease in the level of their job efficiency in the school to their leaving the profession.

Nevertheless, these consequences do not end on the personal level of the teachers, because they also influence wider circles. It has been found, among other things, that teacher burnout has a significant impact on the students, beginning with the negative influence on their achievement and behavior through negative impacts of them dropping out from school (see, for example, Dworkin, 1987; LeCompte & Dworkin, 1991). It has also been found that teacher burnout has a negative influence on the educational system on the national level, even to the point of posing a real threat to its stability (see, for example, Farber, 1991). Among the negative outcomes resulting in the high percentage of attrition from the teaching profession, we can identify a decrease in the quality of teachers and employment of inexperienced or unsuitable teachers, a phenomenon that creates instability in schools and impacts, in a dramatic way, the ability to plan and administer the educational system (Sperling, 2015).

Burnout of Teachers and Emotional Labor in Teaching

Results of numerous studies have shown the relationship between burnout and emotional labor in certain professions and vocations. For example, studies undertaken with nurses in hospitals (Celic et al, 2010) and with hotel hospitality workers (Shani, Uriely, Reichel, & Ginsburg, 2014) indicated the existence of significant correlations between the two variables. However, no significant differences were found between emotional labor and burnout in other professions and vocations, such as in the field of care-giving, administration, office work and physical work (Brotheridge & Grandey, 2002). In the field of teaching, on the other hand, only a few researchers have studied the relationship between emotional labor and burnout. Levine Brown (2011), who reviewed the literature that deals with this relationship among teachers, found that until the time of publication of her study, there had been very few quantitative studies on the topic (Brotheridge & Grandey, 2002; Näring, Vlerick, & van de Ven, 2012) and they all used extremely small samples (less than 40 research participants). In fact, she was the first researcher to examine this question with quantitative tools and with large samples.

Levine Brown (2011) asserts that due to the daily and dynamic interaction between teachers and their students, researchers found that teachers experienced an increase in their use of emotional labor. This increased use of emotional labor could conceivably create burnout, which includes symptoms of emotional exhaustion and exaggerated use of punishment in the classroom to achieve discipline (Näring, Canisius, & Brouwers, 2011). Moreover, Levine Brown assumes that, based on the studies she reviewed, it would be logical to assert that it is possible that teacher burnout is derived from the fact that teachers are unaware that their emotional labor is an inherent part of their work. This assumption receives support from the conclusions of Diefendorff & Gosserand (2003), who noted that by informing teachers about this phenomenon, and by training them in rules related to the demonstration of emotions and emotional labor, their levels of emotional labor and burnout can be moderated.

Studies that examined the relationship between emotional labor and teacher burnout found that it is possible to delay or even prevent burnout by encouraging deep acting by teachers (Carson et al., 2011). In contrast, results of another study carried out among mental health professionals; indicate that the prevention of burnout is brought about by encouraging emotional labor characterized by natural acting (Grandey et al., 2012). In a study carried out in China among college teachers (Zhang & Zhu, 2008), the researchers found that deep acting predicted burnout among; however, the burnout was less than when teachers used surface acting.

The present study aims to examine the relationship between emotional labor and teacher burnout specifically among Israeli female Haredi teachers.

Teachers' Educational Level

Results of research studies have indicated that the higher the teachers' level of education, the higher their sense of self-efficacy (Hoy & Woolfolk, 1993; Cantrell, Young, & Moore, 2003; Akbari & Moradkhan, 2009). Moreover, in a study that examined teacher burnout in Israel, it was found that teachers who only possess a teaching diploma exhibited higher levels of burnout than teachers who held a master's degree (Zaretsky, 2016). In another study, Özan & Şener (2014) examined, among other things, the relationship between the level of education of the teachers and their use of different techniques of emotional labor. These researchers found that teachers who have a master's degree tend to use the natural method of emotional labor as opposed to teachers who only hold a teaching diploma who use the other two techniques associated with emotional labor.

Alavinia & Kurosh (2012) indicated that teachers who maintained higher levels of knowledge and proficiency were characterized by higher levels of teaching efficacy and lower levels of teacher burnout and stress. Swanson and Huff (2010) confirmed that teachers' command of their specific teaching subject and proficiency in their teaching were related to lower levels of teacher attrition and burnout. An increase in the level of academic education makes it possible for the teacher to not only advance economically, but also to achieve promotion to a variety of different positions.

Rationale of the Study

Based on the literature review presented above, which indicated the relevance of emotional labor to teaching and its relationship to other factors, specifically teacher burnout, the present study examined the expressions of emotional labor in teaching among female Israeli Haredi teachers, using a quantitative approach. As can be seen from the articles reviewed above, the topic has been widely studied in other countries: in the United States (see Levine Brown, 2011; Levine Brown et al., 2014), Western Europe (Holland - see Näring et al., 2006; 2011;

2012), Eastern Europe (Poland – see Wróbel, 2013), Asia - Turkey (see Çukur, 2009; Özan & Şener, 2014), Taiwan (see Ching-Sheue, 2015), Pakistan (see Itlaf & Gulzar, 2013), and numerous others. In contrast, in the Israeli context, except for the studies undertaken by Ginat (2011) and Oplatka (2011), very few studies have researched the theory of emotional labor among teachers. One of the main innovations of the present study is that it uses a specially adapted instrument in the Hebrew language for the examination of emotional labor in teaching.

Goal of the Study

The present study aimed to identify the patterns of use of the three techniques of emotional labor (surface acting, deep acting and natural acting) among female Israeli Haredi teachers. In addition, the relationship between burnout level and level of education of the research participants and the emotional labor techniques used by them was examined.

Research Hypotheses

1. Teachers characterized by lower levels of burnout will use the natural acting technique associated with emotional labor whereas teachers typified by higher levels of burnout will utilizing surface acting or deep acting emotional labor techniques.
2. Teachers with higher levels of education will tend to use the natural acting technique associated with emotional labor in contrast to teachers with lower levels of education who will tend to utilize the surface acting or deep acting emotional labor techniques.

Methodology

Participants

The research sample included 170 Haredi (religiously ultra-orthodox) women who served as teachers in elementary and high school throughout Israel. The sample included women of whom (88.8%) were born in Israel, 29/4% of them were in possession of a teaching diploma, 47/6% held bachelor's degrees and 23% held master's degrees. Fifty-seven percent of the teachers were employed in elementary schools and the rest in high schools. Mean age of the teachers was 32.03 years (SD 8.77) and the mean number of years of teaching seniority was 10.61 years (SD 8.28). The decision to investigate a homogenous sample, in terms of gender, was based on methodological considerations of neutralizing the possible influence of the gender variable on the research results.

Instruments

In this study, we used three questionnaires:

1. A demographic questionnaire that was specially compiled for this study.
2. The Emotional Labor of Teaching Scale – TELTS (Levine Brown, 2011) which was translated into Hebrew. This questionnaire, which examined the level of emotional labor of the participants, includes 11 items on a Likert type scale ranging from 1 (low) to 5 (high). The scale reflects three sub-scales that examine the level of use of surface acting (reliability: $\alpha=.80$), deep acting (reliability: $\alpha=.58$), and natural acting (reliability: $\alpha=.65$). For each research participant, the score on each of the three scales was calculated and served as the mean score of participant's responses. A higher score reflected a higher level of use of emotional labor in teaching.
3. The Teacher Burnout Questionnaire (Friedman, 1999) examines the burnout of teachers and is based on Maslach, Jackson & Leiter's (1996) original instrument. The questionnaire comprises 14 items on Likert type scale ranging from 1 (low) to 6 (high) and reflects one general factor (teacher burnout) with reliability of the entire questionnaire being $\alpha=.89$. The mean score for each participant on the burnout factor was based on all responses, with a higher score reflecting a higher level of burnout.

Procedure

Data collection was undertaken via internet-based responses to the research questionnaires. The link to the questionnaires was sent via e-mail to the teachers who participated in the study. In all, 170 full questionnaires were received.

Results

Teacher Burnout and Emotional Labor

To examine any possible differences between teachers characterized by low levels of burnout and teachers typified by high levels of burnout in relation to the use of emotional labor in teaching, we performed an ANOVA statistical procedure. Table 1 presents the means and the standard deviations of the participants' scores for this first hypothesis, namely: Teachers characterized by lower levels of burnout will use the natural acting technique associated with emotional labor whereas teachers typified by higher levels of burnout will utilizing surface acting or deep acting emotional labor techniques.

Table 1. Means and Standard Deviations for Research Participants' Scores on Emotional Labor Factors and Levels of Burnout (N = 170)

Factor	Level of burnout	N	Mean	SD
Surface acting	Low-level burnout	92	2.85	.78
	High-level burnout	78	3.29	.70
Natural acting	Low-level burnout	92	3.96	.58
	High-level burnout	78	3.69	.55
Deep acting	Low-level burnout	92	3.31	.83
	High-level burnout	78	3.38	.64

Results of the analysis indicated significant differences between teachers with a low level of burnout and teachers with a high burnout level, in relation to the technique of surface acting [$F(1,168) = 14.182$, $p<.001$, $\eta^2=0.078$]. Teachers, who exhibited a high level of burnout, were found to be significantly higher in their use of the surface acting technique as opposed to teachers with a low level of burnout. Furthermore, significant differences were found between teachers with a low level of burnout and the teachers with a high level of burnout in relation to their use of natural acting [$F(1,168) = 9.245$, $p<.01$, $\eta^2=0.052$]. Teachers who exhibited a low level of burnout were found to make more use of this technique. We did not find significant differences between teachers with a low level of burnout and teachers with a high level of burnout regarding the use of deep acting [$F(1,168) = 0.279$, $p>.05$, $\eta^2=0.002$].

Educational Level of the Teacher and Emotional Labor

In order to examine second hypothesis, namely teachers with higher levels of education will tend to use the natural acting technique associated with emotional labor in contrast to teachers with lower levels of education who will tend to utilize the surface acting or deep acting emotional labor techniques, potential differences among teachers with different levels of education, and their use of emotional labor in teaching, were analysed by a one-way MANOVA procedure. Results showed that there was a significant difference according to level of education as presented in the following table.

Table 2. Means and Standard Deviations of the Participants' Scores on Emotional Labor Factors and Levels of Education (N = 170)

Factor	Level of Education	N	Mean	SD
Surface acting	Teaching Diploma	50	3.32	.76
	Bachelor's Degree	81	2.99	.77
	Master's Degree	39	2.85	.72
Natural acting	Teaching Diploma	50	3.71	.54
	Bachelor's Degree	81	3.87	.63
	Master's Degree	39	3.91	.52
Deep acting	Teaching Diploma	50	3.45	.71
	Bachelor's Degree	81	3.41	.81
	Master's Degree	39	3.07	.61

Significant findings were found between levels of education and level of use of the surface acting technique [F (2,167) =4.68, p<.05, η^2=0.053]. To ascertain the source of the differences, we performed a post-hoc Scheffe test. The results indicated a significant difference between teachers who only held a teaching diploma and teachers with a master's degree in relation to their use of surface acting. Teachers in possession of a teaching diploma used surface acting significantly more than teachers with a master's degree. Furthermore, significant differences were found between teachers' levels of education and their levels of use of the emotional labor technique of deep acting [F (2,167) =3.38, p<.05, η^2=0.039]. To examine the source of the differences, we conducted a post-hoc Scheffe test. Results indicated that there was a significant difference between teachers who held a master's degree and teachers who possessed a teaching diploma in the use of deep acting. Teachers with a teaching diploma were found to be significantly more prone to adopt the deep acting technique than teachers with a master's degree. In contrast to these results, no significant results were found between teachers' level of education and levels of use of the natural acting technique [F (2,167) =1.65, p>.05, η^2=0.019]. However, a nonsignificant trend indicates that teachers with a master's degree, and to a lesser extent, teachers with a bachelor's degree, used the natural acting technique associated with emotional labor more than teachers who held a teaching diploma.

Discussion

This study examined the relationship between level of teacher burnout and level of teacher education on the one hand and the use of emotional labor techniques on the other. The first hypothesis was that there would be differences between teachers characterized by different levels of burnout and their use of the different techniques associated with emotional labor. The first hypothesis was confirmed, and significant differences were indicated between teachers with different levels of burnout and the use of surface acting. Teachers who had high burnout level were found to use the surface acting technique significantly more than teachers characterized by a low burnout level. These results support the findings of other studies that indicated a positive correlation between burnout symptoms and use of surface acting in general (Brotheridge & Grandey, 2002; Cox, 2012). More specifically, these findings confirm the results of previous studies that indicated a positive relationship between a high level of burnout and the use of surface acting (Zhang & Zhu, 2008; Yoga, Alaskan, & Aru, 2012).

Results of the statistical analysis of the data also indicated that teachers with a low level of burnout tended to use natural acting significantly more than those with a high burnout level. This technique is unique since it is not actually acting, but rather an expression of authentic and naturally honest emotions and feelings. Therefore, some of the studies term this emotional labor technique as the use of genuine emotions, rather than natural acting (Yilmaz et al., 2015). Apparently, workers who express genuine emotions and feelings do not

experience the emotional dissonance that workers experience when they feel compelled to use surface acting or deep acting to conform with conventional norms (see Holman, Chissick, & Totterdell, 2002). This finding also confirms the research results indicated by Yilmaz et al. (2015) who found a negative correlation between burnout level and the expression of genuine emotions and feelings.

The second research hypothesis suggested that there would be differences between teachers with different educational levels and their use of emotional labor techniques in teaching. Statistical analysis of the results indicated that the use of surface acting was significantly more frequent among teachers who held teaching diplomas in comparison with teachers who held master's degrees. These findings support the conclusions of Özan & Şener (2014) who examined the relationship between teachers' educational levels and the emotional labor techniques that they preferred. In addition, the findings of the present study indicate that teachers with a teaching diploma used deep acting significantly more than teachers with either a bachelor's or master's degree. These findings support the results of the study undertaken by Özan & Şener (2014).

The results also indicated a non-significant tendency for teachers with higher educational levels (bachelor's and master's degrees) to use the natural acting technique more frequently than teachers with a teaching diploma, thereby confirming previous findings reported by Hoy & Woolfolk (1993), Cantrell et al. (2003) and by Akbari & Moradkhan (2009). This tendency of teachers with a higher educational level to use the natural acting technique confirms the findings of Yin (2016) who indicated that teachers who use natural acting to genuinely express their emotions in teaching are better able to competently fulfill their professional goals.

In conclusion it may be said that the results of the present study confirm the importance of emotional labor vis-à-vis the performance of teachers. Natural acting is significantly associated with less teacher burnout and stress, thereby enabling the teacher to get on with his or her job without a feeling of dissonance or frustration when facing students. In addition, the results clearly indicate a definite tendency of teachers with more advanced educational levels to use surface acting significantly less than teachers with lower levels of education.

Limitations of the Research

The sample was comprised solely of Haredi women teachers. Further research in Israel should address a more heterogeneous research sample that includes teachers of both genders as well as teachers from the other sectors in Israeli society.

Recommendations for Future Research

This study was a quantitative study. We recommend undertaking additional

studies that adopt a mixed-methods approach. Combining quantitative with qualitative methods can enrich the knowledge that we gain, making it possible to examine the research issues more deeply and comprehensively and to expand the study's scope and possible conclusions.

The present study focused on a specific research population of teachers. This population did not include other types of education professionals, such as administrators, subject coordinators, educational counsellors, special education teachers etc. As seen from the literature review, the need to deal with emotional labor is found in diverse field of service provision. Thus, future research should include other members of the educational staff who fulfil roles other than that of direct teaching.

Practical Recommendations for the Educational System

The first practical recommendation relates to the assimilation among administrators, educational counsellors and teachers of the understanding that emotional labor is not something to be taken for granted, but rather a phenomenon that requires effort that can have a positive influence on the educationalist's feelings and resulting work efficiency.

Following Diekfendorff & Grossrand (2003), and Levine Brown (2011), who indicated that knowledge, practice and assimilation of rules connected to the expression of emotion among teachers moderates the levels of emotional labor of the teachers and the levels of burnout, the importance in relating to the issues of emotional labor in education in different systems becomes clearer. It is vital to address these issues both in pre-service training and in in-service training of teachers in schools (Avdor, 2008). It is especially important to deal with the topic in in-service workshops for new teachers, etc. These recommendations stem from the need to minimize the level of teacher burnout as well as to increase teacher efficiency and effectiveness. This could impact not only teachers, as individuals, but also possibly contribute to the solving of problems on the school and system levels: Issues such as teacher attrition, high levels of turnover, lack of motivation, decrease in the level of efficiency, decrease in the level of commitment to the organization, and more, need to be more proactively addressed for the benefit of the educational system.

Acknowledgement

References

Akbari, R., & Moradkhan, Sh. (2009). Iranian English teachers' self-efficacy: Do academic degrees and experience make a difference? *Pazhuhesh-e Zabanha-ye Khareji*, *56*, 25-47. [Farsi].

Alavinia, P., & Kurosh, S. (2012). On the would-be bonds between emotional intelligence and self-efficacy: The case of Iranian EFL university professors. *Theory and Practice in Language Studies, 2*(5), 956-964.

Ashforth, B. E., & Humphrey, R. H. (1993). Emotional labor in service roles: The influence of identity. *Academy of Management Review, 18*(1), 88-115.

Avdor, S. (2008). *Professional development of teachers and the place of academic training institutions as partners in these processes: A literature review and research findings from the world and Israel.* Authority for the Training of Teaching Professionals in the Ministry of Education, Jerusalem. Retrieved from http://bit.ly/2sIKobY. [Hebrew].

Basim, H. N., Begenirbaş, M., & Yalçin, R. C. (2013). Effects of teacher personalities on emotional exhaustion: Mediation role of emotional labor. *Educational Science: Theory and Practice, 13*(3), 1488-1496.

Brennan, K. (2006). The managed teacher: Emotional labor, education, and technology. *Educational Insights, 10*(2), 55-65.

Brotheridge, C. M., & Grandey, A. A. (2002). Emotional labor and burnout: Comparing two perspectives of "People Work". *Journal of Vocational Behavior, 60*, 17-39.

Cantrell, P., Young, S., & Moore, A. (2003). Factors Affecting Science Teaching Efficacy of Preservice Elementary Teachers. *Journal of Science Education*, *14*(3), 177-192.

Carson, R. L., Plemmons, S., Tenplin, T. J., & Weiss, H. M. (2011). You are who you are: A mixed-method study of affectivity and emotion regulation in curbing teacher burnout. In G. M. Reevy & E. Feydenberg (Eds), *Personality, Stress and Coping: Implications for Education*, (pp. 239-265). Charlotte, NC: IAP Information Age.

Celic, M., Tabak, A., Uysal, M.P., Sigri, U., & Turunc, O. (2010). The relationship between burnout and emotional labor of the employees in hospital sector. *International Journal of Business and Management Studies, 2*(1), 47-54.

Ching-Sheue, F. U. (2015). The effect of emotional labor on job involvement in preschool teachers: Verifying the mediating effect of psychological capital. *Turkish Online Journal of Educational Technology, 14*(3), 145.

Cox, M. E. (2012). Explorations into early care and education providers' job dissatisfaction and mental well-being: Expanding the reach of emotional labor. *Dissertation Abstracts International Section A: Humanities and Social Sciences, 72*(9-A), 3524.

Çukur, C. Ş. (2009). The development of the teacher Emotional Labor Scale (TELS): Validity and reliability. *Educational Sciences: Theory & Practice, 9*(2), 559-574.

Diefendorff, J. M., & Gosserand, R. H. (2003). Understanding the emotional labor process: A control theory perspective. *Journal of Organizational Behavior,* 24(8), 945-959.

Dworkin, A. G. (1987). *Teacher Burnout in the Public Schools: Structural Causes and Consequences for Children*. Albany, NY: SUNY Press.

Ekman, P. & Friesen, W. V. (1975).*Unmasking the face: A guide to recognizing emotions from facial expressions*. Englewood Cliffs, NJ: Prentice Hall.

Farber, B. A. (1991). *Crisis in Education: Stress and Burnout in the American Teacher*.

San Francisco, CA: Jossey-Bass.

Freudenberg, H., J. (1974). Staff burnout. *Journal of Social Issues, 30*(1), 159-165.

Friedman, I. (1999). *Teacher burnout: The concept and its measurement.* Jerusalem: Henrietta Szold Institute.

Gavish, B., & Friedman, I. (2010). Novice teacher's experience of teaching: A dynamic aspect of burnout. *Social Psychology of Education, 13*, 141-167.

Ginat, K. (2011). The researchers of manifest and latent emotions. *Hed Chinuch, 86*(1), 42-45. Retrieved from http://bit.ly/2FijgTN. [Hebrew].

Grandey, A.A. (2000). Emotional regulation in the workplace: A new way to conceptualize emotional labor. *Journal of Occupational Health Psychology*, *5*(1), 95-110.

Grandey, A. A. (2015). Smiling for a wage: What emotional labor teaches us about emotion regulation. *Psychological Inquiry, 26*(1), 54-60.

Grandey, A. A., Foo, S. C., Groth, M., & Goodwin, R. E. (2012). Free to be you and me: A climate of authentic alleviates burnout from emotional labor. *Journal of Occupational Health Psychology, 17*(1), 1-14.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271-293.

Hagenauer, G., & Volet, S. (2014). 'I don't think I could, you know, just teach without any emotion': Exploring the nature and origin of university teachers' emotions. *Research Papers in Education, 29*(2), 240-262.

Hargreaves, A. (1998). The emotional politics of teaching and teacher development: With implications for educational leadership. *International Journal of Leadership in Education: Theory and Practice, 1*(4), 315-336.

Hochschild, A. R. (1983). *The managed heart: Commercialization of human feeling*. Berkeley, CA: University of California Press,.

Hochschild, A.R. (2012). *The managed heart: Commercialization of human feeling* (updated with a new preface). Berkeley and Los Angeles, CA: University of California Press.

Holman, D., Chissick, C., & Totterdell, P. (2002). The effects of performance monitoring on emotional labor and well-being in call centers. *Motivation and Emotion, 26*(1), 57-81.

Hoy, W.K., & Woolfolk, A.E. (1993). Teachers' sense of efficacy and the organizational health of schools. *The Elementary School Journal*, *93*(4), 355-372

Iltaf, H., & Gulzar, A. (2013). Impact of expressivity and impulse strength on burnout and turnover intentions: Mediating role of deep acting. *Journal of Business and Management, 7*(4), 62-67.

Larrivee, B. (2012). *Cultivation Teacher Renewal: Guarding against Stress and Burnout.* Lanham, MD: Rowman & Littlefield Education.

LeCompte, M.D., & Dworkin, A.G. (1991). *Giving Up on school: Student Dropouts and Teacher Burnouts*. Newbury Park, CA: Corwin Press.

Levine Brown, E. (2011). Emotion matters: Exploring the emotional labor of teaching. Unpublished Doctoral dissertation, University of Pittsburgh, PA.

Levine Brown, E., Horner, C.G., Kerr, M.M.. & Scanlon, C.L. (2014). United States teachers' emotional labor and professional identities. *KEDI Journal of Educational Policy, 11*(2), 205-225.

Merriam-Webster Online Dictionary. (2017). *Emotion*. Retrieved from http://bit.ly/ 2cLkSdf.

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

Maslach, C., Schaufeli, W.B., & Leiter, M.P. (2001). Job burnout. *Annual Review of*

Psychology, 52(1), 397-1, 397-422.

Matsumoto, D., & Hwang, H. C. (2013). Cultural similarities and differences in emblematic gestures. *Journal of Nonverbal Behavior, 37*, 1-27.

Moran, C. M., Diefendorff, J. M., & Greguras, G. J. (2013). Understanding emotional display rules at work and outside of work: The effects of country and gender. *Motivation and Emotion, 37*(2), 323-334.

Näring, G.W.B., Canisius, A.R.M., & Brouwers, A. (2011). Measuring emotional labor in the classroom: The darker emotions. In A. Caetano., S. Silva, & M.J. Chambel (Eds), *New challenges for a healthy workplace in human services,* (pp. 127-139). Munich: Rainer Hampp Verlag.

Näring, G., Vlerick, P., & van de Ven, B. (2012). Emotion work and emotional exhaustion in teachers: The job and individual perspective. *Educational Studies, 38*(1), 63-72.

Oplatka, Y. (2009). Emotion management and display in teaching some ethical and moral considerations in the era of marketization and commercialization. In P. A. Schutz & M. Zembylas (Eds.), *Advances in teacher emotion research: The impact on teachers' lives,* (pp. 55-71). New York, NY: Springer.

Oplatka, Y. (2011). Service providers or educators? *Hed Hachinuch, 86*(1)*, 46-49. Retrieved from http://bit.ly/2HztrE0. [Hebrew].

Özan, M. B., & Şener, G. (2014). Teaching and emotional labor. *American International Journal of Social Science, 3*(5), 111-119.

Pfister, L. L. (2015). *Emotional display rules for clerical workers, teachers, custodians, and cafeteria workers in Pennsylvania K-12 public school organizations.* Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, PA.

Roberts, K. R. L. (2011). *Emotional labor, emotional expression, and emotional control in the K-12 classroom* Unpublished doctoral dissertation, University of Alabama, Tuscaloosa, AL. Retrieved from http://bit.ly/2HxEMEM.

Sahin, E. (2015). Determination of associations between educational beliefs, emotional labor and self-efficacy in teacher candidates. *Anthropologist, 19*(2), 543- 555.

Shani, A., Uriely, N., Reichel, A., & Ginsburg, L. (2014). Emotional labor in the hospitality industry: The influence of contextual factors. International Journal of Hospitality Management, *37*, 150-158.

Sperling, D. (2015). *Teacher attrition in the world: A review of information.* Mofet, Tel-Aviv: Institute and the Inter-College Center for Information. [Hebrew].

Subic-Wrana, C., Manfred, E. B., Brähler, E., Stöber-Richter, Y., Knebel, A., Lane, R. D., & Wilting, J. (2014). How is emotional awareness related to emotion regulation strategies and self-reported negative affect in the general population? *Plos One, 9*(3), 1-10.

Sun, Y. (2013). *The developmental characteristics of preschool teachers' emotional labor and its relation to emotional exhaustion.* Unpublished doctoral dissertation, Northeast Normal University, Changchun, China.

Sutton, R. E. (2004). Emotional regulation goals and strategies of teachers. *School Psychology of Education, 7*(4), 379-398.

Swanson, P., & Huff, R. (2010). The relationship of Georgia's rural foreign language teachers' sense of efficacy to teacher attrition. *The Rural Educator*, *31*(3), 16-29.

Tsang, K. K. (2014). A review of current sociological research on teachers' emotions: The way forward. *British Journal of Education, Society & Behavioural Science, 4*(2), 241-256.

Wong, E., Tschan, F., Messerly, L., & Semmer, N. K. (2013). Expressing and amplifying positive emotions facilitate goal attainments in workplace interactions. *Frontiers in Psychology, 4*, 188.

Wróbel, M. (2013). Can empathy lead to emotional exhaustion in teachers? The mediating role of emotional labor. *International Journal of Occupational Medicine and Environmental Health, 26*(4), 581-592.

Ye, M., & Chen, Y. (2015). A literature review on teacher's emotional labor. *Creative Education, 6,* 2232-2240.

Yilmaz, K., Altinkurt, Y., Guner, M., & Sen, B. (2015). The relationship between teachers' emotional labor and burnout level. *Eurasian Journal of Educational Research*, *59*, 75-90.

Yozgat, U., Çalışkan, S. C., & Ürü, F. O. (2012). Exploring emotional dissonance: On doing what you feel and feeling what you do. *Procedia-Social and Behavioral Sciences, 58*, 673-682.

Zaretsky, R. (2016). *The theory of emotional labor in teaching: The impact of seniority of the teacher, age of the students and the level of burnout of teachers on their perceptions of the rules connected to demonstration of emotions, and on the use of techniques of emotional labor in teaching*. Unpublished master's thesis, Bar Ilan University, Ramat-Gan, Israel. [Hebrew].

Zhang, Q., & Zhu, W. (2008). Exploring emotion in teaching: Emotional labor, burnout, and satisfaction in Chinese higher education. *Communication Education, 57*(1), 105-122.

Response to "An Examination of Plausible Score Correlation from the Trend in Mathematics and Science Study"

By Plamen Vladkov Mirazchiyski[*]

This article is a response to an article written by Wang and Ma "An Examination of Plausible Score Correlation from the Trend in Mathematics and Science Study", published in the Athens Journal of Education. The purpose of this paper is to address issues with Wang's and Ma's suggestion to use analysis method for correlating plausible values from TIMSS. The paper reviews the design of international large-scale assessments, and TIMSS in particular, and its implications for data analysis in regard to the application of Canonical Correlation Analysis for calculating association between two sets of plausible values, as proposed by Wang and Ma. The conclusion is that, given the design of TIMSS and other large-scale assessments, the method proposed by Wang and Ma is not appropriate for correlating two sets of plausible values because they are not multivariate measures as the suggested method would assume. Some other methodological issues related to the overall analysis approach used by Wang and Ma is discussed as well.

Keywords: correlation, international large-scale assessments, methodology, multiple imputation, plausible values

Introduction

This paper is a response article that addresses some issues with an article by Wang and Ma (2016) published in the *Athens Journal of Education*. The original article published by the aforementioned authors represents an attempt to address methodological issues when correlating two sets of plausible values (PVs) from the Trends in International Mathematics and Science Study (TIMSS) that the authors claim to exist.

The main argument of the authors is that if two sets of PVs are used in correlation analyses of TIMSS data with the current approach the study uses, this can inflate the chance for making Type I error due to the non-additive nature of the correlation coefficients. The solution of this assumed issue the authors suggest is Canonical Correlation Analyses (CCA) that can accommodate for the multidimensionality of the signals and avoid the dependency on the coordinate system in which the variables are described (Wang & Ma, 2016). However, most of the assumptions the authors have do not reflect the design and the methodology TIMSS, as well as other large-scale assessments (ILSA), as well as the analytical strategy and methods it uses to produce unbiased estimates from any statistical analysis being employed.

---

[*]Chief Executive Officer, Researcher, International Educational Research and Evaluation Institute (INERI), Slovenia.

The next section provides a review of the literature on generation of PV sin general and as they are used in TIMSS (as well as in other ILSA), what is the necessity for their derivation as proficiency scores, what they are, how they are derived and how they shall be used in analysis reflecting their nature and statistical theory that stands behind their derivation. In addition, the literature review presents the sampling procedures and the derivation of the sampling weights which must be included in the computation of any estimates using ILSA data. This is necessary because Wang and Ma (2016) do not address the important issues of the unequal probabilities of selection and the subsequent issue of sampling variance when computing estimates with TIMSS data. The literature review ends with an overview of the CCA, its assumptions and application in situations where other correlation methods are not appropriate. The purpose of this review is to add more clarity on the design of TIMSS and ILSA in general, and thus, justify the current approaches in using them in analysis. These provide the background against which the claims in Wang's and Ma's (2016) article will be discussed in the last, discussion, section.

Literature Review

TIMSS is the successor of other ILSA conducted prior 1990s: the First International Mathematics Study (or FIMS, conducted in 1964), the First International Science Study (or FISS, conducted in 1970-1971). These studies, conducted by the International Association for the Evaluation of Educational Achievement (IEA) had a follow-up cycles in the period 1980-1984, the Second International Mathematics Study (SIMS) and the Second International Science Study (SISS) (IEA, n.d.). TIMSS, initially named as Third International Mathematics and Science Study, was the third cycle of both mathematics and science studies where they were conducted jointly (Mullis et al., 1997). In 1999 the IEA replicated TIMSS and named it as TIMSS-R or TIMSS-Repeat (Mullis et al., 2000), and due to the decision of conducting the study in regular cycles, later "third" was changed to "trends in".

TIMSS, as well as other studies, used the methodological developments originating in the National Assessment of Educational Progress (NAEP) conducted in US, extending these advancements (Rutkowski, Gonzalez, Joncas, & von Davier, 2010). Due to student fatigue, attrition and logistics of ILSA there was a need to find a solution to make possible carrying out assessment on a large scale to cope with the aforementioned limitations – not everyone can be tested using every single item. One such solution was the use of Multiple Matrix Sampling (MMS). MMS, used in many studies, is different than subjects or examinees sampling where the subjects (usually students) are selected from the population of interest. In MMS the measures on which the subjects are tested or surveyed on are sampled from a universe of interest, i.e. part (a sample) of the total assessment (Rutkowski, Gonzalez, von Davier, & Zhou, 2014). This facilitates the testing of the sampled subjects in broad content domains like mathematics or science where a large number of items are needed to have a reliable measure in the domain of

interest. In addition, TIMSS has several content sub-domains in both mathematics and science, but also cognitive sub domains in both areas. This would lead to estimated testing time as of more than 10 hours in TIMSS 2007, for example (Rutkowski et al., 2014). Thus, in TIMSS 2007 the total of 429 mathematics and science items were distributed in 14 blocks, each containing unique set of items, in each content domain (mathematics and science), rotated across 14 test booklets, each containing two mathematics and two science blocks. One of the mathematics blocks and one of the science blocks in each booklet was repeated in every next booklet, so that a link through common blocks across the booklets was ensured. This way, the testing time and logistic demands of the entire study are decreased, making it possible to conduct such a large study across a myriad of countries at the same time (Rutkowski et al., 2014). This way, no tested subject is taking every single item, but a sample of items which removes the burden from the examinees and is a cost-effective solution. Early developments of this technique have proven that group means appear to be more consistent than a sample of tested subjects taking all items. Currently there are different implementations of the MMS, one of them, also used in TIMSS is the Block Incomplete Booklet (BIB) set of designs (Rutkowski et al., 2014).

The use of MMS saves a lot of efforts and minimizes the testing time for the students participating in ILSA. However, this brings a serious challenge when estimating the student proficiency. Following the description of the MMS provided above,

> The relatively small number of items per block and the relatively small number of blocks per test booklet mean that the accuracy of measurement at the individual level of these assessments is considerably lower than is the level of accuracy common for individual tests used for diagnosis, tracking, and/or admission purposes(von Davier, Gonzalez, & Mislevy, 2009, p. 11).

Traditional methods for estimating the proficiency of tested subjects would yields "biased or inconsistent variance estimates of population parameters" (Rutkowski et al., 2010, p. 145). As stated previously, none of the sampled student takes all items, but (as for TIMSS 2007 example given above) two mathematics blocks and two science blocks containing unique items that do not appear in any other block; the students answer only the items presented to them and for the rest of the items student answers are missing by design. However, items differ in their characteristics, difficulty being the most important. Hence, percent correct for the items a student faced will not be the appropriate method because it limits comparability of results – the score will depend on the particular set of items a particular student receives (Mirazchiyski, 2013). The use of Item Response Theory (IRT) was more and more needed due to the use of MMS in assessments (Rutkowski et al., 2014). The traditional IRT approaches that use Marginal Maximum Likelihood (MML) and Expected a Posteriori (EP), however, are not appropriate solutions as well. These estimation techniques produce point estimates optimized for individual-level, but not group level estimation (von Davier et al., 2009). Several scholars developed group level models for estimating latent traits

stemming from measurements using MMS. NAEP uses these models since its assessment in 1983/84 to find a tractable solution for estimating the standard errors. The use of population models and their applications in cases of MMS together with IRT is commonly referred to "plausible values", although often other names are used as well (Rutkowski et al., 2014). The derivation of PVs from population models relies on Rubin's multiple imputation methods developed in the period between late 1970s to late 1980s. These models impute the tested subjects' scores for references on population level (Rutkowski et al., 2014).

Latent traits (such as intelligence or reading skills among many other) are not directly observable (von Davier, 2014) and in ILSA (and not only) are treated as missing (Rutkowski et al., 2010). Instead observing the latent traits directly, it is possible to observe the responses of examinees to tasks they face as indicators of these under lying traits. It is necessary to know, however, how these observable indicator variables relate to the latent trait. ILSA use IRT involving latent regression of the latent (unobservable) proficiency variable on number of predictor variables. This approach includes the information of all observable variables and follows the models of imputing data developed in the 1980s and 1990s. This means that the conditional distribution of the latent variable depends on the values of the observed variables, assuming that the missing values are missing at random (von Davier, 2014). As von Davier points out, "100% of the student proficiency data is imputed using a specialized imputation model based on statistical procedures that are tailored to incorporate both cognitive response data and student background data" (von Davier, 2014, p. 184). ILSA utilize latent regression models which provide Expected a Posteriori (EAP) estimates of posterior variance of the measured ability. These latent regression models are actually an extension of the multiple group IRT model. What they provide is a "different conditional prior distribution for each respondent's proficiency based on a set of predictor variables" (von Davier, 2014, p. 184). Although the total number of achievement items in each ILSA is large, it is still limited for each student due to MMS. Thus, it is complemented with the items from the background questionnaires, applying Rubin's multiple imputation approach to impute the answers of the items the student did not face and create student ability distribution for the entire population or sub-populations of interest (Rutkowski et al., 2010). It may not be immediately obvious, but PVs "add exactly the right amount of variability to make the distribution of the PVs in the group match the distribution of the true values in the group" (von Davier et al., 2009, p. 35). The foundations of the PVs methodology, its theoretical rationale, foundations and mathematical proof are laid out in Mislevy (1991) and Mislevy, Beaton, Kaplan and Sheehan (1992). An overview of what PVs are for the non-technical reader is provided by von Davier et al. (2009).

When it comes to the actual application of the population modeling and imputing the missing data along with the latent regression models used in ILSA, each study has its own specifics, although many things in common as well. The presentation here continues with a description of the TIMSS 1995 proficiency scaling methodology because this is the study and cycle the authors of the original article (Wang & Ma, 2016) used. The subsequent cycles of TIMSS use the same approach and steps for scaling the cognitive data, although some details may

differ.

In TIMSS 1995 first a subsample of 600 students is drawn from each country, forming an "international calibration samples" that are about equal in number of selected students. These samples were drawn systematically with Probability Proportional to the Size (PPS) of the Primary Selection Units (PSU) (i.e. schools) using their overall weights as measure of size. This led to equal selection probabilities within the national samples to draw the calibration samples within each country, thus each country was given an equal weight in estimating the item parameters in the next step(Adams, Wu, & Macaskill, 1997). The scaling model applied was a generalization of the more basic unidimensional model. In addition, a multivariate linear model imposed on the population distribution. The item parameters were estimated using the international sample. Then the model was fit in each country using fixed item parameters obtained in the previous step (Adams et al., 1997). The population model uses the item response model which is a conditional model describing the process of generating the responses conditional on the latent variable. The derivation of conditioning variables is done using background information to form a vector included in the latent regression model as a predictor. Then the five PVs are derived for each student by making five random draws from the formed marginal posterior latent distribution (Adams et al., 1997). The chapter on the scaling methodology and procedures in the TIMSS 1995 Technical Report (Adams et al., 1997) provides more comprehensive technical and detailed information. The corresponding chapter in the TIMSS 2007 Technical Report (Foy, Galia, & Li, 2008) provides more reader-friendly description of the scaling procedures.

The important detail from the review of the methodology and procedures of obtaining the PVs as proficiency scores that has to be stressed is that a set of PVs for each subject area (e.g. overall science) or content sub-domain (e.g. chemistry) or cognitive sub-domain (e.g. applying) is a set of variables representing unidimensional measure of the same construct of interest. As explained above, a set of PVs represent five random draws from a marginal posterior latent distribution. That is, each one of the PVs in a set (e.g. overall science score) represents a measure on the same construct that includes information from the same cognitive items and the same background variables that are used to construct the marginal posterior distribution they are drawn from.

As stated earlier, the procedures of obtaining PVs follow the theory of multiple imputation using the information from the background variables as predictors in a latent regression model. Rubin (1987) provides a theoretically and methodologically sound background on the imputation techniques, models and analysis with the resulting imputed data sets. As Mislevy (1993) notes, Rubin's approach of imputing missing data multiple times creates data sets where "each missing variable is replaced by a draw from its predictive distribution, conditional on the observed data" (Mislevy, 1993, p. 79). The generation of PVs follows the same logic, as already described in the previous paragraphs. The methods for analysis of imputed data sets were specified and complemented in subsequent publications (e.g. Little & Rubin, 1987, 2002). The five randomly drawn PVs for each student vary in their values as a result of the multiple imputation. When it

comes to analysis of PVs, five estimates of any statistics are computed with each of the five PVs (or any measure that has been imputed multiple times) and they are all different. This is a result of what is called "imputation variance" or "imputation error" (Foy et al., 2008) which reflects the measurement error stemming from the use of MMS (Foy et al., 2008). All analyses that include PVs have to follow the approach of performing analyses with multiple imputation variables. Rubin (1987) and (Little & Rubin, 1987, 2002) provide a set of rules in order to combine the parameter estimates and compute the variance associated with the multiple imputation of measures. These rules of combining the estimates and compute the imputation variance have found strong empirical support in various studies as in Schafer (1999) and in the case of PVs in various papers, as in von Davier et al. (2009) and Rutkowski et al. (Rutkowski et al., 2010). Following the theoretical developments of Rubin (1987) and Little an Rubin (1987, 2002), any analysis of TIMSS 1995 involving PVs will perform the computations five times (once with each PV) and the results of these computations will be averaged to obtain an unbiased estimate of student performance (Gonzalez, 1997). Formula for computing the imputation variance in TIMSS and ILSA in general are provided by Foy et al. (2008) and (von Davier et al., 2009), to name just few, and the technical report of each ILSA provides such formulas reflecting the specifics of the study. The same approach is used not only in TIMSS, but in other ILSA, such as PIRLS, ICCS, ICILS and PISA, and are implemented in statistical software which will be discussed later in the paper.

The imputation variance, however, is not the only source of error in ILSA. Besides the sample of items from the universe of all possible items that can measure given construct, ILSA also use sample of students from the target population for which the construct is measured. The sampling design of TIMSS 1995 is a two-stage stratified cluster sampling design and is done separately for each population of interest. In the first stage in each participating country 150 schools where students in a particular population of interest study are sampled with PPS. Schools are the PSUs. The second stage of sampling picks intact classrooms within the sampled schools (clusters). Usually one intact classroom is sampled, although some participated countries preferred two. In some countries a third stage (sampling students within the selected classrooms) was added, but these were exceptional cases. Due to the clustering effect of selecting intact classrooms (students in the same classroom tend to be more alike), intraclass correlation (ICC, a measure of similarity within a cluster) and the size of the classroom, along with the desired standard error from the sample, were taken into account when calculating the desired sample size within each country (Foy, Rust, & Schleicher, 1996). A stratification (grouping of schools according an attribute) was applied in most countries to improve sampling efficiency, making estimates more reliable, to apply different sample design to specific groups of schools or regions, and to ensure adequate representation of specific groups in the target population. Explicit stratification would construct independent lists of schools on an attribute. Implicit stratification would use the same list of schools where schools are sorted by the attribute (Foy et al., 1996). In the first stage of sampling, schools in the sampling frame (or frames, in case of explicit stratification) are sorted by their measure of

size (MOS) (number of students in the target population). A sampling interval is defined by dividing MOS by the number of schools to be sampled. The first school is selected by choosing a random number between 1 and the number representing the sampling interval. Thus, the number obtained represents the MOS of the first school being selected. Adding the sampling interval to this number would give the number of the second sampled school, and so on. If an implicit stratification is applied, the sampling of schools will reflect the implicit strata within which the schools are sorted by their MOS too. At the second sampling stage one or two (depending on the countries preferences) intact classes were sampled (Foy et al., 1996).

Not only TIMSS, but all other ILSA as well, follow the same or similar sampling strategy. This kind of sampling (PPS) is rather different than the Simple Random Sampling (SRS). The difference is that with SRS every student is selected with probability that is equal to the probability of selecting any other student (Rutkowski et al., 2010). On the contrary, as presented above, in TIMSS (as well as all other ILSA), do not use SRS. The application of PPS sampling means that each sampled school, hence classroom and student within it, are sampled with different probability that depends on the number of students in the target grade attending the schools. Then different students in the sample will not represent the same number of students in the population they were selected from. In addition, different countries chose different stratification variables in the sampling process to satisfy their research demands (Foy, 1997). This is an additional challenge for analyzing data that stems from TIMSS or other ILSA. Sampling weights in ILSA are introduced to accommodate for the sampling with different probabilities to ensure that certain groups in the population of interests are not overrepresented in the sample (Rutkowski et al., 2010). The TIMSS 1995 weighs calculation was done in three steps. First, calculating the school weights, adjusting for the school non-response independently for each design domain or explicit stratum. Second, calculating the classroom weight adjusting for the non-response of the classroom. When only one classroom was sampled, no classroom adjustment was necessary. Third, computing the sampling weight of the participating students adjusting for their non-response. The final weight for each student was added as product of the three intermediate weights from the previous steps. The weights computed in the first three steps before their non-response adjustments are computed as the inverse probabilities of selection (for the school, class and student) (Foy, 1997).

The TIMSS, or any other ILSA, sampling design provides country samples in ILSA representative for the population they have been drawn from (Foy, 1997). Not using the sampling weights in analyzing ILSA data leads to giving more importance on some students due to the sampling design. Relevant example in this regard is provided by Rutkowski et al. (2010) who demonstrate how the sampling of students from different school types biases the results on population level when weights are not used to adjust for the number of students in the population each one of the sampled students represents.

An important issue in ILSA is the variance estimation to compute the standard errors due to the stratified multi-stage sampling. The standard procedures and

formulas for computing the standard errors do not apply to ILSA because the sampling strategy does not rely on SRS. The IEA and OECD studies rely on replication techniques to estimate the variance. TIMSS, in particular, uses Jackknife Repeated Replication (JRR). When analyzing ILSA data, replication techniques should be used for all sampling variability estimates, which are the sampling errors, to obtain unbiased estimates (Rutkowski et al., 2010). The necessity of using replication methods for sampling variance estimation stems from the sampling strategy that uses unequal probability to sample schools teaching students in a target population to obtain efficient and cost-effective samples. The JRR variation in TIMSS 1995 assumes that PSUs (schools) can be paired according to the sampling design, forming pseudo-stratums (pairs of schools) to estimate the sampling variance. This approach appropriately reflects the combined effect of the within- and between-school contribution to the sampling variance (Gonzalez & Foy, 1997). The procedure is as follow. First, sampling zones (paired schools) are formed. The sampling zones are formed in the same order in which the schools were sampled. With 150 sampled schools, 75 zones are formed. When more schools are sampled, sometimes schools were combined before forming a sampling zone. Second, the variance is estimated in each sampling zone by setting the weight of one of the paired schools to 0 and doubling the weight of the other school. The estimation is done 75 times plus once with the full weight. At the end, the variance is estimated by combining the results. More details and the formula for combining the estimates to produce the sampling variation of an estimate is provided by Gonzalez and Foy (1997). This procedure is rather different than standard methods of estimating the error under SRS. Rutkowski et al. (2010) provide a clear example what are the consequences when replication is not applied using TIMSS 2007 data.

When using PVs, each estimate is computed five times (once with each PV) and within each JRR zone. The standard error of an estimate using PVs is computed using both the sampling and imputation variance components. Both of these components are important, omitting any of them can produce biased result. Pedagogical examples are provided by Rutkowski et al. (2010). Formulas for combining the estimated imputation and sampling variance, as well as the total standard error of an estimate, are provided for each ILSA reflecting the specifics of the study and even the study's cycle. Such can be found in Foy et al. (2008) and Schulz (2011), for example.

The systematic publications on correlating two sets of variables begin in 1936 with a publication of Hotelling (1936), although some elements were developed earlier by Bravais, Galton, Pearson, Yule and others (Hotelling, 1936). Hotelling (1936) was concerned with issues in correlating two sets of variables representing multidimensional measures and suggests the name "canonical correlations". Later, this kind of analysis was referred to mainly as Canonical Correlation Analysis (CCA). Borga (2001), also cited by Wang and Ma (2016), defines CCA as "a way of measuring the linear relationship between two multidimensional variables" (Borga, 2001, p. 2), that is, each set of variables would represent different measures, directly observable or not. Härdle and Simar (2007) define it as technique for analyzing the association of two data sets based on projections where

"an index (projected multivariate variable) that maximally correlates with the index of the other variable" (Härdle & Simar, 2007, p. 321). Similarly, Borga (2001) defines the CCA as "finding two sets of basis vectors, one for **x** and the other for **y**, such that the correlations between the projections of the variables onto these basis vectors are mutually maximized" (Borga, 2001, p. 2). Thus, CCA "is based on linear indices, i.e., linear combinations of the random variables" (Härdle & Simar, 2007, p. 321). The most important thing to note about CCA is that it is a method of correlating two multidimensional sets of variables, i.e. the variables in each set are measures that quantify different properties of the objects. The next section will reiterate this when discussing the CCA as a method for correlating PVs in TIMSS, suggested by Wang and Ma (2016).

Discussion

Wang and Ma (2016) raised concerns about the calculation of correlation coefficients as computed in TIMSS when two sets of PVs are used. The concerns were raised due to the non-additive nature of correlation coefficients and suggested the use of canonical correlation instead to reduce the risk of making Type I error. However, as already clarified in the literature review, a set of Plausible Values (PVs) does not contain multiple different measures on multiple different latent traits as CCA would assume. The important details that needs to be reiterated here is that set of five PVs in any subject area, content or cognitive subdomain is derived as five random draws from the same marginal posterior distribution (Adams et al., 1997), i.e. each PV in a set of PVs (e.g. overall mathematics) represents the same measure, carrying out the information obtained using the same items, their IRT parameters, conditioned on the same background variables. Hence, it is the same trait presented as five different variables (PVs), and not five different latent traits in five different variables. Being imputed variables, all rules that apply to analysis of multiply imputed data sets (see Little & Rubin, 2002) apply when working with PVs regardless of the analysis type. As Mislevy (1993) notes, this situation resembles the situation of having multiple unbiased and conditionally independent indicator variables, but only on the surface. Further, he demonstrates the pitfalls of using PVs as unbiased and conditionally independent indicators of a latent variable, producing incorrect estimates. Given the answer of the third research question Wang and Ma (2016) have, the correct approach for correlating the two sets of mathematics and science achievement PVs would be to correlate the first PV in mathematics with the first PV in science, the second PV in mathematics with the second PV in science, and so on, then averaging the obtained estimates to derive the final estimate of the correlation. The computation of the sampling and imputation variance and the final standard error of this coefficient would follow formulas which can be found in Foy et al. (2008). This kind of pairing will not allow underestimation of the correlation between the subjects. There are software products that are capable to do this tedious work with minimal effort from the side of the analyst. One such product is the IDB Analyzer, freely available from the IEA (IEA, 2016) where the correlation of two sets of PVs

follows the routine described above. The standard error, in turn, is necessary to test the statistical significance of the estimates. Another issue that deserves attention in regard to the sampling variance as component of the standard error is that there is no indication if the authors used any weights. As mentioned above, the sampling weights must be used to properly estimate the sampling variance. This issue is discussed further.

The aforementioned issue is the main issue with Wang's and Ma's (2016) publication: CCA is not the appropriate method for analyzing ILSA data stemming from any study that uses the PVs methodology because each PV represents the same unidimensional latent trait, and not multiple dimensions of a latent trait or multiple traits as CCA assumes. Information on what PVs are and how to use them in analysis of student achievement, along with the literature sources was already provided by the literature review.

There are other points of concern with the Wang's and Ma's (2016) article, as outlined below.

First, on page 305 the authors provide a citation from Garcia (2010) saying that "One cannot add raw r values to compute an arithmetic average an r" (Garcia, 2010, p. 2), but missed important information located few lines below where the author of the original publication adds that,

> "in order to compute $\bar{r}$ [that is, average correlation] individual r values have to be converted into additive quantities. Several techniques, each with their own assumptions and drawbacks, can be used: transcendental transformations, numerical expansions, weighted averages, or combination of techniques" (Garcia, 2010, p. 2).

Further, the authors added a citation from Statsoft (2000) to reconfirm that,

> "Because the value of the correlation coefficient is not a linear function of the magnitude of the relation between the variables, correlation coefficients cannot simply be averaged" (StatSoft, 2000, p. 10).

It is very strange, however, that the authors missed the next sentence that says that,

> "In cases when you need to average correlations, they first have to be converted into additive measures. For example, before averaging, you can square them to obtain coefficients of determination which are additive (as explained before in this section), or convert them into so-called Fisher z values, which are also additive." (StatSoft, 2000, p. 10).

There is one important thing to note in the above citations, as well as with the sources as a whole: they consider a general case and do not pertain to analyzing multiple imputation data sets.

Second, on page 307 the authors wrote that "After completing canonical correlation analyses, the results are merged with mathematics and science performance scores at the country level to address Question 3. The combined data

set is attached in Appendix 1."

Unfortunately, they do not provide any detail on how the CCA results were computed and merged to the TIMSS 1995 mathematics and science scores. The methodology of extracting information from variables to add it later to the same variables shall be explained: 1) why was it done, what was the purpose; and 2) how was it done. Also, besides the data, the aforementioned appendix should contain combined data and tables with the canonical correlations, as mentioned on page 309. However, there is no appendix to this article and these are not possible to inspect.

Third, on page 307 the authors wrote that "The inclusion of canonical correlation as a predictor automatically assumes co-existence of mathematics achievement as an explanatory variable". It is quite unclear what is meant by this, especially how a statistical method can be included as a predictor. Probably they meant that the results of CCA were included as a predictor. However, as mentioned above, there is no information how the CCA results were merged with the scores.

Fourth, on page 307 the authors added SPSS syntax to compute regression analysis, noting that it is a "simple SPSS application without involvement of complex Macro syntax on the variable dimension" (Wang & Ma, 2016, p. 309). As mentioned above, it is not clear how the canonical correlations were computed and included in the analysis this syntax uses. The provided syntax is not only simple, it is oversimplified and does not take into account any of the design issues TIMSS 1995 has, as the JACKREGPV. SPSmacro provided by Foy, Arora and Stanco (2013) and recommended by Statistics Canada (2002) does. The syntax provided by Wang and Ma (2016) does not include any statement that weights the data using any of the weighting variables available in the TIMSS 1995 data and do not use the variance estimation methods as used in TIMSS (see Foy et al., 2008) to estimate the sampling variance and take it into account when computing the standard errors. In the discussion part of the paper, on page 309, the authors mention for the first time that the TIMSS uses multistage sampling. TIMSS, as well as other ILSA, uses sampling with probability proportional to the size of the schools. Thus, schools, and their students respectively, are sampled with different probabilities. To be able to produce estimates on population level, which is the purpose of TIMSS and any other ILSA, weights that account for the unequal probability of selection must be used. Due to the PPS sampling, if no weights are used, some students can have a disproportional impact on the estimates and the analysis can provide rather biased results, as demonstrated by Rutkowski et al. (2010). An additional issue is that with large samples, as in TIMSS, not using appropriate methods of variance estimation will underestimate the SE. This is probably the case with the strong and significant results the authors obtained in their study, not using any weight and their replication in the analysis. Given all of the above, the conclusion the authors make on page 309 that "the influence from complex sampling is washed out" (Wang & Ma, 2016, p. 309) along with the arguments is incorrect simply because weighting matters for all types of statistical estimates with ILSA data, including TIMSS. As per the use of design effect itself for estimating the sampling variance, it is largely discouraged in the recent years.

There are different methods to compute the standard errors due to the clustering effect of the sampling used in large-scale assessments, one of them being the design effect. It is an estimate of how large the effects of dependency among observations are, including clustering, on the sampling variance. Despite the ease of calculating design effects, they appear as rather crude estimates of the standard errors which actually appear to be inflated (Barron, 2000) and conservative compared to other methods (Stapleton, 2008). In addition, an empirical study conducted by Stapleton (2008) concludes that design effect are only appropriate for univariate statistics, when applied accurately. In contrast, using jackknifing to compute standard errors is much more precise (Barron, 2000). As Barron (2000) notes, the computation of jackknife estimates of standard errors has been difficult for secondary analysts. However, in recent years computers have become more powerful and software for using jackknifing is available.

Conclusions

The method for correlating sets of PVs suggested by Wang and Ma (2016) is an attempt to solve an alleged problem in analyzing TIMSS 1995 data. However, as this paper demonstrates, the authors do not take into account the assessment design of TIMSS 1995. This resulted in recommendation of analysis method that has assumptions and applications not relevant to analysis of PVs due to their specifics. Further, the actual application of the method ignores another important design issue, the complex sampling design of TIMSS 1995, and the necessary use of sampling weights. This is proven to have negative effects when conducting analyses with ILSA data regardless of the analysis method.

The contemporary ILSA are tools for policy making in education. The decisions made from analysis results have an impact on the on the implementation of policies and reforms in education. It is a great responsibility of researchers using these data to apply appropriate analysis methods, taking into consideration the study design and nature of the measures. Otherwise, biased results presented to policy makers may lead to ineffective policies.

References

Adams, R. J., Wu, M. L., & Macaskill, G. (1997). Scaling Methodology and Procedures for the Mathematics and Science Scales. In M. O. Martin & D. L. Kelly (Eds.), *Third International Mathematics and Science Study: Technical Report* Vol. 2, pp. 111–146. Chestnut Hill, MA: Boston College.

Barron, S. (2000). Difficulties Associated with Secondary Analysis of NAEP Data. In N. S. Raju, J. W. Pellegrino, M. W. Bertenthal, K. J. Mitchell, L. R. Jones, & National Research Council (U.S.) (Eds.), *Grading the Nation's Report Card: Research from the Evaluation of NAEP*, pp. 172–194. Washington, D.C: National Academy Press.

Borga, M. (2001, January 12). Canonical Correlation: A Tutorial. Retrieved from https:// bit.ly/2OfnsX0.

Foy, P. (1997). Calculation of Sampling Weights. In M. O. Martin & D. L. Kelly (Eds.),

Third International Mathematics and Science Study: Technical Report Vol. 2, pp. 71–79.

Foy, P., Arora, A., & Stanco, G. M. (Eds.). (2013). *TIMSS 2011 User Guide for the International Database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College and the IEA.

Foy, P., Galia, J., & Li, I. (2008). Scaling the Data from the TIMSS 2007 Mathematics and Science Assessments. In J. F. Olson, M. O. Martin, & I. V. Mullis (Eds.), *TIMSS 2007 Technical Report* , pp. 225–280. Chestnut Hill, MA: Boston College.

Foy, P., Rust, K., & Schleicher, A. (1996). Sample Design. In M. O. Martin & D. L. Kelly (Eds.), Vol. 1, pp. 4-1-4–17. Chestnut Hill, MA: Boston College.

Garcia, E. (2010). *A tutorial on correlation coefficients*. Retrieved from https://bit.ly/2Q6 SYbC.

Gonzalez, E. J. (1997). Reporting Student Achievement in Mathematics and Science. In M. O. Martin & D. L. Kelly (Eds.), *Third International Mathematics and Science Study: Technical Report*, Vol. 2, pp. 147–174. Chestnunt Hill, MA: Boston College.

Gonzalez, E. J., & Foy, P. (1997). Estimation of Sampling Variability, Design Effects, and Effective Sample Sizes. In M. O. Martin & D. L. Kelly (Eds.), *Third International Mathematics and Science Study: Technical Report*, Vol. 2, pp. 81–100. Chestnut Hill, MA: Boston College.

Härdle, W., & Simar, L. (2007). *Applied multivariate statistical analysis* (2nd ed.). Berlin: Springer.

Hotelling, H. (1936). Relations between Two Sets of Variates. *Biometrika*, *28*(3/4), 321–377.

IEA. (2016). *IEA IDB Analyzer* (Version 4.0) [Computer software]. Hamburg, Germany: IEA.

IEA. (n.d.). *Other IEA studies.* Retrieved January 3, 2018, from https://bit.ly/2OeApRa.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

Mirazchiyski, P. (2013). *Providing School-Level Reports from International Large-Scale Assessments: Methodological Considerations, Limitations, and Possible Solutions*. Amsterdam: International Association for the Evaluation of Educational Achievement.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, (56), 177–196.

Mislevy, R. J. (1993). Should "Multiple Imputations" be Treated as "Multiple Indicators"? *Psychometrika*, *58*(1), 79–85.

Mislevy, R. J., Beaton, A. E., Kaplan, B., & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*(2), 133–161.

Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1997). *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: TIMSS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., Gregory, K. D., Garden, R. A., O'Connor, K. M., … Smith, T. A. (2000). *TIMSS 1999 International Mathematics Rreport: Findings from IEA's Repeat of the Third International Mathematics and Science Study at the Eighth Grade*. Chestnut Hill, MA: International Study Center, Boston College, Lynch School of Education.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.

Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting. *Educational Researcher*, *39*(2), 142–151.

Rutkowski, L., Gonzalez, E., von Davier, M., & Zhou, Y. (2014). Assessment Design for International Large-Scale Assessments. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of International Large-Scale Assessments: Background, Technical Issues, and Methods of Data Analysis*, pp. 75–96. Boca Raton, FL: CRC Press.

Schafer, J. L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*, *8*, 3–15.

Schulz, W. (2011). The reporting of ICCS results. In W. Schulz, J. Ainley, & J. Fraillon (Eds.), *ICCS 2009 Technical Report*. Amsterdam: International Association for the Evaluation of Educational Achievement.

Stapleton, L. M. (2008). Analysis of data from complex surveys. In E. D. Leeuw, J. J. Hox, & D. A. Dillman (Eds.), *International Handbook of Survey Methodology*, pp. 342–369. New York, NY: Lawrence Erlbaum Associates.

Statistis Canada. (2002). *The International Adult Literacy and Skills Survey, 2003*. Ottawa, ON: Statistics Canada.

StatSoft. (2000). Basic Statistics. Retrieved August 20, 2017, from https://bit.ly/2AyXA 4Q.

von Davier, M. (2014). Imputing Proficiency Data under Planned Missingness in Population Models. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of International Large-Scale Assessments: Background, Technical Issues, and Methods of Data Analysis*, pp. 175–202. Boca Raton, FL: CRC Press.

von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series*, *2*(1), 9–36.

Wang, J., & Ma, X. (2016). An Examination of Plausible Score Correlation from the Trend in Mathematics and Science Study. *Athens Journal of Education*, *3*(4), 301–312.

Rejoinder: Response to: "An Examination of Plausible Score Correlation from the Trend in Mathematics and Science Study"

By Jianjun Wang[*] & Xin Ma[†]

This rejoinder keeps the original focus on statistical computing pertaining to the correlation of student achievement between mathematics and science from the Trend in Mathematics and Science Study (TIMSS). Albeit the availability of student performance data in TIMSS and the emphasis of the inter-subject connection in the Next Generation Science Standards in the U.S., the major TIMSS reports did not consistently include the correlational findings of student achievements, nor did the TIMSS macro, JACKREGPV per Mirazchiyski's recommendation, meet the need of correlation computing between each pair of plausible values (PV) in mathematics and science. In addition, an inflation of Type I error would be inevitable unless a feasible approach is taken to concurrently correlate the two sets of PVs from each subject. In this context, we adduced canonical correlation analysis (CCA) as a feasible approach that included considerations of both sampling weight and standard error adjustment due to complex sampling. More importantly, this rejoinder reveals two fundamental issues in the TIMSS literature: (1) Incompliance to the multiple imputation rule in reporting the TIMSS 1995 international findings (see Beaton et al., 1996a, b); (2) Need of substantially increasing the number of PV imputations as illustrated by the mainstream statistical computing in general (e.g., SAS Institute, 2015; STATA, 2017), and the recent international large scale assessments (ILSA) in particular.

Keywords: International Studies, TIMSS Data Analysis

It is our great honor to contribute this rejoinder to an article of our esteemed colleague, Plamen Vladkov Mirazchiyski, the former Deputy Head of the Research and Analysis Unit of the International Association for the Evaluation of Educational Achievement (IEA). We appreciate his effort on writing an extensive response (we shall brief his article name as "response") that nearly doubled the space of our original article. We sincerely accept his criticism on the briefness of our report that was originally designed for a 20-minute presentation at the ninth Annual International Conference on Statistics (Athens, Greece) in June, 2015.

In this rejoinder, we are given the opportunity to elaborate our research with a clear focus on the potential disagreements we seemed to have with Mirazchiyski. To keep it simple, we categorized this article into four parts. First, we will analyze the core issues that originated this track of in-depth inquiry. In the second section, we clarify what is not the issue. The third section is devoted to coverage of new literature beyond the foundation of multiple imputation method Mirazchiyski

---

[*]Professor, California State University, Bakersfield, USA.
[†]Professor, Department of Educational, School, and Counseling Psychology, University of Kentucky, USA.

reviewed from the period of 1970s – 1990s. We conclude this rejoinder in the last section to support the momentum of knowledge advancement in comparative education.

What Is the Issue?

Mirazchiyski raised issues on several fronts, including (1) the incorporation of sampling weights and the adjustment of result variance due to stratified, cluster sampling, (2) the ignorance of using TIMSS tools, particularly the macro of JACKREGPV.SPS, in the statistical computing that involves plausible values, (3) the use of design effect that, albeit its candidacy as a viable method for handling data with complex sampling in many decades (Kish, 1965), might produce crude estimates of the standard errors, (4) the inappropriateness of using the canonical correlation method in our original work for computing correlation coefficients.

All these issues appear to be handled with a simplistic question on what works. While it is still legitimate, a more precise question should be on "what works in which context." In reviewing the research context of the Trend in Mathematics and Science Study (TIMSS), we appreciate Mirazchiyski's attention on other [international] large scale assessments (ILSA)[1], such as the First International Mathematics Study (or FIMS, conducted in 1964), the First International Science Study (or FISS, conducted in 1970-1971), the Second International Mathematics Study (SIMS), and the Second International Science Study (SISS) in the 1980s. He also noted the National Assessment of Educational Progress (NAEP) conducted in the United States, which is not an international study. However, it was extended in the two phases of International Assessment of Educational Progress (IAEP) prior to TIMSS. Further, the NAEP method was largely adopted by TIMSS.

Throughout his response, Mirazchiyski did not pay any attention to the context so as to contrast TIMSS with other studies. One special feature of TIMSS is the opportunity of analyzing the correlation of student performance between mathematics and science that never existed in the past ILSA. Although FIMS, FISS, SIMS, SISS, and IAEP covered the subjects of mathematics and science, these assessments were conducted on different student samples, and thus, no attempt would be appropriate to compute correlation coefficients between mathematics and science scores.

Indeed, the inter-subject connection is important. For instance, the Next Generation Science Standards (NGSS) indicated strong needs to connect mathematics and science education in the United States (National Research Council, 2014). Similar inquiry-based interdisciplinary learning was emphasized in the previous version of the U.S. curriculum standards since the mid 1990s. TIMSS is unique and unprecedented for surveying mathematics and science performance from the same group of students concurrently.

Despite its 20-year history from TIMSS 1995 to TIMSS 2015, major TIMSS

---

[1]We took the liberty to add "International" because it was missing in Dr. Mirazchiyski's acronym definition.

reports are divided according to the subject boundary, one for mathematics (Beaton et al., 1996a; Mullis et al., 2000; Mullis, Martin, Gonzalez, & Chrostowski, 2004; Mullis, I.V.S., Martin, M.O., & Foy, 2008; Mullis, Martin, Foy, & Arora, 2012a; Mullis, Martin, Foy, & Hooper, 2016a) and the other for science (Beaton et al., 1996b; Martin et al., 2000; Martin, Mullis, Gonzalez, & Chrostowski, 2004; Martin, Mullis, & Foy, 2008; Mullis, Martin, Foy, & Arora, 2012b; Mullis, Martin, Foy, & Hooper, 2016b), as if there were no connections between them.

More importantly, TIMSS methods, particularly the JACKREGPV.SPS macro adduced by Mirazchiyski, are not prepared for supporting the correlation analyses of student achievement between mathematics and science. Because TIMSS did not produce a macro for correlation computing, it seems natural to first obtain the coefficient of determination (R^2) from a simple regression that contains one dependent variable and one independent variable. The correlation coefficient can be subsequently computed from the square root of the R^2 result. In his Discussion section, Mirazchiyski dictated,

> [T]he correct approach for correlating the two sets of mathematics and science achievement PVs would be to correlate the first PV in mathematics with the first PV in science, the second PV in mathematics with the second PV in science, and so on, then averaging the obtained estimates to derive the final estimate of the correlation. (p. 14)

which concurred the paring approach of using one plausible value (PV) from mathematics and another PV from science. For five PVs in each subject, five rounds of repetitions are needed prior the result aggregation.

Nonetheless, to invoke the JACKREGPV.SPS macro from the TIMSS tools, researchers are expected to concurrently use all five plausible values (see NPV=5 below), instead of one PV from each subject[2]. As Foy, Arora, and Stanco (2013) acknowledged, "[t]he JACKREGP macro [sic] is used to perform a multiple linear regression between *a set of plausible values as the dependent variables* [italics added for emphasis] and a set of independent variables" (p. 52), which does not work for paring the PV for correlation computing in Mirazchiyski's afore-quoted suggestion.

---

[2]The screenshot here comes from page 163 of https://bit.ly/1sFVLw9.

```
Include "c:\jackregpv.sps".

Jackregpv
                Infile    = temp              /
                Cvar     = cntrid             /
                Xvar     = regsex             /
                Rootpv = Prose                /
                NPV      = 5                   /
                Njkz      = 30                 /
                Rpwt     = replic01 to replic30 /
                Wgt       = popwt.
```

We further question Mirazchiyski's claim of support for JACKREGPV or JACKREGP [sic] application from Foy, Arora, and Stanco (2013) and Statistics Canada (2002). Although TIMSS has made a convention of using the macro with PV in the name for analyses that involve plausible values (Foy, Arora, & Stanco, 2013), the macro that can be used for the simple regression analysis with one PV as the dependent variable is JACKREG.SPS. In other words, the convention of Foy, Arora, and Stanco (2013) does not apply here when we handle correlation analyses with a pair of PVs, instead of all five PVs from both sides of mathematics and science. For that reason, the computation cannot be supported by the PV macros, such as JACKPV and JACKREGPV, from TIMSS.

In summary, the main issue seemed to hinge on the potential oversight of TIMSS colleagues on the unique opportunity to correlate student performance between mathematics and science that never occurred in previous ILSA projects. Although 20 years have lapsed between TIMSS 1995 and TIMSS 2015, no major reports were disseminated on the correlation outcomes, nor did the TIMSS team clarify the confusion on which macro to revoke for the computing. By choice, we picked TIMSS 1995 data in our study to help track the ongoing disengagement of TIMSS reporting on the correlation part since its inception.

What Is Not the Issue?

The previous section addressed the issues that are important to us but was overlooked in Mirazchiyski's response. In this section, we intend to reciprocally discuss topics that have been considered as issues by him, but not us.

First, we agree that TIMSS data analyses need to consider both sampling weight due to unequal rates of sample selection and variance inflation for using a stratified, cluster sample structure. We also agree that canonical correlation analysis (CCA) is a method not solely developed to serve TIMSS researchers. For the sampling weight part, CCA does not have the issue because standard statistical software packages such as SAS allow incorporation of sampling weight in computing the coefficient from canonical correlation[3]. We were curious on why Mirazchiyski argued that our method "ignores another important design issue, the

---

[3] p. 11 of https://bit.ly/2StobHJ.

complex sampling design of TIMSS 1995, and the necessary use of sampling weights" (p. 18).

The adjustment of variance to address the impact of complex sampling also needs to be done. However, it can be considered in multiple ways. For instance, SAS PROC SURVEYMEANS offered both design effect and Jackknife options for researchers to choose because supporters can be found in either aisle[4]. While the variance could be inflated by a factor of k (the design size) due to complex sampling (Kish, 1965), our article also clarified the cancellation of the k factor in correlation computing because of its involvement at both numerator and denominator of the formula. Mirazchiyski did not dispute our special attention for the correlation computing. Instead, he seemed to come from the Jackknife aisles of practitioners and wrote: "As per the use of design effect itself for estimating the sampling variance, it is largely discouraged in the recent years" (p. 17). In terms of the reason, he argued without an independent study that it produced "rather crude estimates of the standard errors" (p. 17). Nonetheless, even with a crude k value estimation as he alleged, little impact could have occurred in our correlation computing due to the numerator and denominator parts of the correlation configuration noted in our article.

> In reviewing the past TIMSS practice, Mirazchiyski acknowledged,
> Following the theoretical developments of Rubin (1987) and Little and Rubin (1987, 2002), any analysis of TIMSS 1995 involving PVs will perform the computations five times (once with each PV) and the results of these computations will be averaged to obtain an unbiased estimate of student performance (Gonzalez, 1997). (p. 8)

Indeed, Rubin's (1987) breakthrough has offered several desirable features, including (1) introducing appropriate random errors beyond any deterministic single imputation and (2) offering concise rules for combining the results from multiple imputations for statistical inference.

Despite Mirazchiyski's respect for Rubin's (1987) requirement on the results aggregation, a single PV was used to represent student achievement score in TIMSS 1995 result reporting (see Beaton et al., 1996a, b). According to Gonzalez and Smith (1997, ch. 6, p. 3), the essential step of Rubin's (1987) was not taken because of a decision to ignore the imputation error. When the imputation error was not considered, what was the purpose of spending the precious resources for multiple imputation (MI)? Arguing that one set of the imputed plausible scores can be considered as good as another (Gonzalez & Smith, 1997, ch. 6, p. 3), the TIMSS report fell back to the single imputation result to generate the findings from the first PV. Consequently, the choice of other imputed plausible scores may result in alternative findings different from those in the released TIMSS reports (Wang, 2001).

Mirazchiyski did hint on the detailed differences between TIMSS 1995 and the subsequent cycles of TIMSS in result reporting:

---

[4]https://bit.ly/2qhV4Tr.

The presentation here continues with a description of the TIMSS 1995 proficiency scaling methodology because this is the study and cycle the authors of the original article (Wang & Ma, 2016) used. The subsequent cycles of TIMSS use the same approach and steps for scaling the cognitive data, although some details may differ. (p. 6)

Without his elaboration on the details of the differed steps, we take the liberty to clarify one difference in the major international reports between TIMSS 1995 and its subsequent waves of trend studies. The reporting issue should be treated seriously according Mirazchiyski's following statement,

The contemporary ILSA are tools for policy making in education. The decisions made from analysis results have an impact on the [on the] implementation of policies and reforms in education. It is a great responsibility of researchers using these data to apply appropriate analysis methods, taking into consideration the study design and nature of the measures. Otherwise, biased results presented to policy makers may lead to ineffective policies. (p. 18)

Despite the widespread negative impact of the TIMSS 1995 reports (Beaton et al., 1996a, b) against modeling the professional practice in MI application, we no longer treat this as a long-lasting issue. Mirazchiyski's following notes show his agreement with us on the PVs as different variables:

The five randomly drawn PVs for each student vary in their values as a result of the multiple imputation. When it comes to analysis of PVs, five estimates of any statistics are computed with each of the five PVs (or any measure that has been imputed multiple times) and they are all different. (p. 8)

More importantly, the incompliance did not repeat itself in other waves of TIMSS reporting (e.g., Mullis et al., 2000; Mullis, Martin, Gonzalez, & Chrostowski, 2004; Mullis, I.V.S., Martin, M.O., & Foy, 2008; Mullis, Martin, Foy, & Arora, 2012a; Mullis, Martin, Foy, & Hooper, 2016a, b).

Although Mirazchiyski agreed on the "imputation variance" or "imputation error" that caused the differences among five PVs, he stressed dimensionality considerations that treat PVs as "a set of variables representing unidimensional measure of the same construct of interest" (p. 7). In particular, Mirazchiyski criticized us for using CCA because "a set of Plausible Values (PVs) does not contain multiple different measures on multiple different latent traits as CCA would assume" (p. 13). Unfortunately, we still do not think this as an issue because CCA does not have the assumption on multiple different latent traits. As Borga (2001) pointed out,

Canonical correlation analysis (CCA) is a way of measuring the linear relationship between two multidimensional variables. It finds two bases, one for each variable that are optimal with respect to correlations and, at the same

time, it finds the corresponding correlations. In other words, it finds the two bases in which the correlation matrix between the variables is diagonal and the correlations on the diagonal are maximized. *The dimensionality of these new bases is equal to or less than the smallest dimensionality of the two variables* [italics added for emphasis]. (p. 2)

In theory, if the five PVs have a latent dimensionality equal to 1 for either science or mathematics performance, CCA can still generate statistical findings on these new bases of the latent traits with unidimensionality. In practice, the feasibility was reconfirmed by CCA computing in SAS without any warning or error messages pertaining to the dimensionality checking. Built on the relation between CCA and factor analysis, Levine (1977) also confirmed that just as one (general) factor can occur in factor analysis, so can unidimensionality happen in CCA.

In summary, we agree with Mirazchiyski on the importance of incorporating sampling weight to ensure that certain groups are not overrepresented in the sample for each education system in ILSA. We also agree that additional consideration should be given to the adjustment of variability index due to complex sampling. These methods, including the use of design effect and Jackknife approaches, are included in standard software such as SAS for survey data analyses. Further, the procedure for CCA was well-established in SPSS or SAS to allow dimensionality of the latent bases to range from 1 to another integer, depending on the vector settings. Therefore, we have clarified the issues that seemed important to Mirazchiyski, but not us.

In summary, our original analyses of the TIMSS data were built on in-depth understanding of the sample survey, data imputation, and CCA computing with support from standard software packages in the market. Meanwhile, even with the identification of an indisputable issue in the TIMSS 1995 reports (Beaton et al., 1996a, b) for incompliance to the MI application rule established by Rubin (1987), we still do not treat it as a permanent issue because the compliance problem has been corrected in the other waves of TIMSS reports. From the perspective of supporting ongoing improvement for the benefit of our professional community, we reviewed additional literature below to sustain the momentum of methodology advancement in TIMSS.

What Does the Current Literature Say?

The literature indicates that the employment of Multiple Matrix Sampling design (as in TIMSS) results in "[t]he relatively small number of items per block and the relatively small number of blocks per test booklet" (von Davier, Gonzalez, & Mislevy, 2009, p. 11). Because the total number of achievement items in each ILSA is large, the task of missing value imputation is not small. The imputation model also involves information from background questionnaire to compose predictors. Mirazchiyski relied on the past practice to acknowledge that "[t]he derivation of PVs from population models relies on Rubin's multiple imputation

methods developed in the period between late 1970s to late 1980s" (p. 5).

At that time, Rubin (1987) convinced the research community that five imputations were adequate for treating missing data in general. Software packages, such as SPSS and SAS, accepted this convention to set the number of default imputation to 5 in their MI procedures. Allison (2001) applauded that "Like maximum likelihood, multiple imputation estimates are consistent and asymptotically normal. They are close to being asymptotically efficient" (p. 81).

More recently, Bodner (2008) illustrated that important quantities (e.g., p values, confidence intervals, and estimated fractions of missing information) suffer from substantial imprecision with a small number of imputations. Sullivan, Salter, Ryan, and Lee (2015) concurred that "increasing the number of imputations entails greater precision" (p. 553). Practices of MI computing since the 1990s have led to an increase on the number of imputations in major standard software packages. For instance, SAS Institute (2015) announced that "the default number of imputations in PROC MI has been changed from NIMPUTE=5 to NIMPUTE=25 in SAS/STAT 14.1" (p. 5921). The STATA software manual also indicated that "we recommend using at least 20 imputations to reduce the sampling error due to imputations" (STATA, 2017, p. 5).

Like TIMSS, the MI method implemented by SAS, SPSS, and STATA are all grounded on Rubin's (1987) classic work. However, Wang and Johnson (2018) compared the results of SPSS and SAS computing when the number of imputations was kept at 5. They found that a predictor variable could be claimed both significant and not significant depending on the software being used. As Von Hippel (2016) noted, "[n]on-replicable results reduce scientific openness and transparency, and the possibility of changing results by re-imputing the data offers researchers an opportunity to capitalize on chance by imputing and re-imputing the data until a desired result, such as p<.05, is obtained" (p. 2).

More critically, when more variables are involved in the model setting and a large portion of missing information exists in the data set, researchers may encounter a high risk of obtaining conflicting MI results because the number of imputations is set too small in computer-based analyses (Wang & Johnson, 2018). To overcome this kind of problems, Allison (2012) recommended that "if 27% of the cases in your data set have missing data on one or more variables in your model, you should generate about 30 imputed data sets" (p. 1). Hence, after more than 20 years since the completion of TIMSS 1995, the research literature calls into question the optimism of the TIMSS guideline to delimit the number of PV imputations at 5. Finally, we note that PISA (Program for International Student Assessment), another important ILSA project, has already increased the number of PVs from 5 to 10 in its recent cycles[5].

If more PVs were imputed according to the latest wisdom of the research community, a much larger number of PVs would be released in the TIMSS database. Consequently, the risk of Type I error would substantially increase in computing correlation coefficients from more pairs of PVs between mathematics and science. Although this inflation of Type I error is a totally separate issue from

---

[5]https://bit.ly/2qjBs8B.

the non-additive nature of the correlation coefficients, Mirazchiyski insisted that "The main argument of the authors [i.e., Wang & Ma, 2016] is that if two sets of PVs are used in correlation analyses of TIMSS data with the current approach the study uses, this can inflate the chance for making Type I error *due to the non-additive nature of the correlation coefficients* [italics added for emphasis]" (p. 1-2).

In retrospect, this section is built on the latest research literature to urge for imputing more PVs in TIMSS and make it conform to the current professional practice in statistical computing. Meanwhile, more exploration, such as CCA in our original paper, should be encouraged to help control Type I error under the new circumstances. As the number of PVs is on the rise in current ILSA, the CCA approach may offer more advantage to correlate multiple outcome measures for the benefit of computational efficiency and control of Type I error.

Conclusion

TIMSS produces rich comparative information, and it is a daunting task to comprehensively document correct analytical methods for all data analysts. Nonetheless, colleagues in charge of TIMSS data dissemination confronted the challenge by providing a user guide for secondary data analysis (e.g., Foy, Arora, & Stanco, 2013). It should not come as a surprise that researchers find some analytical circumstances that are not clearly articulated by the TIMSS user guide. For instance, researchers are advised to perform computations five times (once for each PV) and average the results (Gonzalez, 1997). The average mechanism, per instruction of Little and Rubin (2002), was an arithmetic average without referring to any transformations. TIMSS researchers never cited StatSoft (2000, p. 10) on the need for converting correlation coefficients into additive measures. We are glad that Mirazchiyski followed a citation from our original article (i.e., StatSoft, 2000) to amend the part not covered by the TIMSS user guide. As we deal with the ongoing knowledge development, there is no need for assuming panacea on what worked. Instead, more attention should be given on what works in a specific context.

Accommodating more imputations is another front that imposes technical challenges for TIMSS data producers to catch up with the recent advancement in statistical computing. The existing tools, including macros such as JACKREGPV and the IDB Analyzer (IEA, 2016), are not flexible for supporting correlation of student performance between mathematics and science with an adequate number of imputations much larger than 5. While we have completely addressed our rationale for using CCA and design effect, TIMSS researchers should be allowed to continue using the methods of their choice, such as Jackknife computing to adjust standard errors that are not delimited to the correlation analyses. Furthermore, ongoing exploration is needed to control the inflation of Type I error should the number of imputations be increased to a default of 25, as accomplished by a standard software package like SAS.

Finally, we wish to acknowledge that Mirazchiyski' response demonstrated

his rich knowledge about TIMSS practices. The entire research community owes him and his colleagues for their persistent effort in producing large-scale international data that are important in comparative studies. In terms of the value of TIMSS, our rejoinder might have inadvertently made his responses like preaching to the converted. Still, his article clearly offered a great opportunity for clarifying the details of TIMSS methodology. It has been our privilege to interact with this exemplary scholar in the field of ILSA.

Acknowledgement

References

Allison, P. (2001). *Missing data*. Retrieved from https://bit.ly/2AyhVHG.

Allison, P. (2012). *Why you probably need more imputations than you think*. Retrieved from https://bit.ly/2Q30mot.

Beaton, A.E., Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Kelly, D.L., & Smith, T.A. (1996a). *Mathematics achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Beaton, A.E., Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Smith, T.A., & Kelly, D.L. (1996b). *Science achievement in the middle school years: IEA's Third International Mathematics and Science Study (TIMSS)*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Bodner, T.E. (2008). What improves with increased missing data imputations? *Structural Equation Modeling, 15*, 651-675.

Borga, M. (2001, January 12). Canonical correlation: A tutorial. Retrieved from https://bit.ly/1XWgwB9

Foy, P., Arora, A., & Stanco, G.M. (1 Eds.) (2013). *TIMSS 2011 user guide for the international database*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College and the IEA.

Gonzalez, E.J. (1997). Reporting student achievement in mathematics and science. In M.O. Martin & D.L. Kelly (Eds.), *Third International Mathematics and Science Study: Technical report,* vol. 2, pp. 147–174. Chestnut Hill, MA: Boston College.

Gonzalez, E.J., & Smith, T.A. (1997). *Users guide for the TIMSS international database*. Chestnut Hill, MA: TIMSS International Study Center.

IEA. (2016). IEA IDB Analyzer (Version 4.0) [Computer software]. Hamburg, Germany: IEA.

Levine, M.S. (1977). *Canonical analysis and factor comparison*. Newbury Park, CA: Sage.

Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons.

Little, R.J.A., & Rubin, D.B. (2002). *Statistical analysis with missing data* (2nd ed.). Hoboken, NJ: Wiley.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., Gregory, K.D., Smith, T.A., Chrostowski, S.J., Garden, R.A., & O'Connor, K.M. (2000). *TIMSS 1999 international science*

report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., & Chrostowski, S.J. (2004). *TIMSS 2003 international science report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Martin, M.O., Mullis, I.V.S., & Foy, P. (with Olson, J.F., Erberber, E., Preuschoff, C., & Galia, J.). (2008). TIMSS 2007 International Science Report: Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., Gregory, K.D., Garden, R.A., O'Connor, K.M., Chrostowski, S.J., & Smith, T.A. (2000). *TIMSS 1999 international mathematics report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., & Chrostowski, S.J. (2004). *TIMSS 2003 international mathematics report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., & Foy, P. (with Olson, J.F., Preuschoff, C., Erberber, E., Arora, A., & Galia, J.). (2008). TIMSS 2007 international mathematics report. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012a). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Foy, P., & Arora, A. (2012b). *TIMSS 2011 international results in science*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I.V.S., Martin, M.O., Foy, P., & Hooper, M. (2016a). *TIMSS 2015 international results in mathematics*. Retrieved from https://bit.ly/2nAWGOk.

Mullis, I.V.S., Martin, M.O., Foy, P., & Hooper, M. (2016b). *TIMSS 2015 international results in science*. Retrieved from https://bit.ly/2PtZcoX.

National Research Council (2014). *Developing assessments for the Next Generation Science Standards*. Washington, D.C.: The National Academies Press.

Rubin, D. B. (1987) Multiple imputation for nonresponse in surveys. New York: Wiley.

SAS Institute (2015). *SAS/STAT 14.1 user's guide*. Gary, NC: Author.

STATA (2017). *Stata multiple imputation reference manual (Release 15)*. College Station, TX: Author.

StatSoft (2000). *Basic statistics*. Retrieved from http://bit.ly/1rpeZFf

Sullivan, T., Salter, A., Ryan, P., & Lee, K. (2015). Bias and precision of the "Multiple Imputation, Then Deletion" method for dealing with missing outcome data. *American Journal of Epidemiology, 182*(6), 528-534. doi:10.1093/aje/kwv100.

Von Hippel, P. (2016). *The number of imputations should increase quadratically with the fraction of missing information.* Retrieved from https://arxiv.org/ftp/arxiv/papers/1608/1608.05406.pdf.

Wang, J. (2001). TIMSS primary and middle school data: Some technical concerns. *Educational Researcher, 30*(6), 17-21.

Wang, J., & Johnson, D. (2018). An examination of discrepancies in multiple imputation procedures between SAS and SPSS. *The American Statistician*. doi: 10.1080/0003 1305.2018.1437078.