

The background of the cover features a close-up photograph of a red ribbon and a silver medal resting on a wooden surface. The ribbon is draped across the top and right sides, while the medal is positioned in the lower-left quadrant. The wood grain is clearly visible, adding a natural and textured feel to the design.

Journal of Physical Education and Sport Management

Volume 5 Number 2 May 2014

ISSN 1996-0794



*Academic
Journals*

ABOUT JPESM

Journal of Physical Education and Sport Management (JPESM) is published twice a month (one volume per year) by Academic Journals.

Journal of Physical Education and Sport Management (JPESM) is an open access journal that provides rapid publication (monthly) of articles in all areas of the subject such as physical fitness ,sports agent laws ,sport organization, health care etc.

The Journal welcomes the submission of manuscripts that meet the general criteria of significance and Scientific excellence. Papers will be published shortly after acceptance. All articles published in JPESM are peer-reviewed.

Contact Us

Editorial Office: jpesm@academicjournals.org

Help Desk: helpdesk@academicjournals.org

Website: <http://www.academicjournals.org/journal/JPESM>

Submit manuscript online <http://ms.academicjournals.me/>.

Editors

Prof. Chung Pak-kwong

*Head, Department of Physical Education
Hong Kong Baptist University,
Hong Kong.*

Prof. Nader Rahnema

*Department of Nutrition, Food
and Exercise Sciences,
424 Sandels Building,
Florida State University,
Tallahassee, FL 32306,
USA.*

Dr. Van Hung

*College of Education, Vietnam
National University, Hanoi
CauGiay Dist, Hanoi,
Vietnam.*

Editorial Board

Nikolaos Digelidis

*Assistant Professor
LLP-Erasmus Departmental Co-ordinator
Member of the Institutional Committee of International
Affairs
University of Thessaly
Department of Physical Education and Sport Science
Karies, GR-42100 Trikala
Greece*

Prof. Zan Gao

*Department of Health, Exercise and Sport Science
Texas Tech University
Box 43011, 105 PrinTech
Lubbock, TX 79409-3011*

Prof. Me. Fabio Scorsolini-Comin

*Department Of Developmental Psychology, Education
and Labor
Institute Of Education, Humanities, Arts, Humanities
And Social Sciences
University Of Delta Mining (Uftm)
Guaritá Vargas Avenue, 159, Abbey
Cep: 38025-440
Uberaba - Mg
Brazil*

Prof. Maycon de Moura Reboredo

*Federal University of Juiz de Fora
Brazil*

Aldo M. Costa *Assistant*

*Professor University of
Beira Interior
Department of Sport Sciences, Covilhã, Portugal
Portugal*

Dr. GV Krishnaveni

*Epidemiology Research Unit
Holdsworth Memorial Hospital
Mysore 570021
Ph: 0821 2529347
India.*

Prof. Sydney Schaefer

*Program in Physical Therapy
Washington University School of Medicine
United States*

Dr. Alexandra Lopes

*University of Porto
Via Panorâmica s/n 4150-564 Porto
Portugal.*

Prof. Daniel A. Marinho

*University of Beira Interior/CIDESD, Covilhã.
Portugal*

Prof. M^a José Martínez Zapata, MD

*Centro Cochrane Iberoamericano-Servicio de
Epidemiologia Clínica
IIB Sant Pau.
Sant Antoni M^a Claret, 171.
Barcelona 08041.*

Prof. Teresa F. Amaral

*Nutritionist
Faculty of Nutrition and Food Sciences,
University of Porto & System Integration and Process
Automation UISPA-IDMEC, University of Porto
Rua Dr. Roberto Frias
4200-465 Porto,
Portugal*

Full Length Research Paper

Using dance to tap into the kinesthetic learning capacities of young English Language learners in the United States

Liane Brouillette

Center for Learning through Arts and Technology 3200 Education Building
University of California, Irvine, CA 92697-5500, USA.

Received 9th September 2011; Accepted 31st August 2012; Published May 2014

This study reports on the kind of school leadership style that best suits for promoting teachers' job satisfaction in primary schools in Tanzania. The study employed cross sectional research design with samples of 200 teachers from 20 selected primary schools in Songea and Morogoro districts. Interviews, documentary analysis and questionnaires were used to collect data. The data were analyzed both qualitatively and quantitatively using tables, frequencies and percentages. It was found that the democratic leadership style was the most dominant in best performing primary schools. It is therefore suggested that there is much to be learnt with democratic leadership style as a copying strategy in least performing primary schools. Moreover, level of teachers' job satisfaction was reportedly high in best performing schools compared to least performing schools. The findings commended that democratic leadership style is the one which promotes high teachers' job satisfaction among teachers in primary schools.

Key words: Leadership, leadership styles, primary schools, teachers and job satisfaction.

INTRODUCTION

This article focuses on the use of dance and creative movement to transcend language barriers in early elementary classrooms where many of the children speak a language other than English at home. Throughout history, dance has been used to educate, celebrate and symbolize. We will explore how creative movement can provide limited English-proficient students with opportunities for vocabulary building and social-emotional development, regardless of their level of language proficiency or cultural background. Dance will be investigated both as a form of artistic literacy and as a

powerful means of interpersonal communication. Globalization has increased the flow of international migration in recent years. Currently, there are approximately 168 million international migrants; about 60 percent of these have moved from developing nations to industrialized nations. This has created challenges for schools in the receiving nations, where children may arrive at school speaking many different languages. This article describes the pressures created by demographic change and looks at how the kinesthetic learning capacities of young children might be used as a cultural bridge. We will

E-mail: lbrouill@uci.edu.

Author agree that this article remain permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

explore how creative movement provides young limited English-proficient students with opportunities for vocabulary building and social-emotional development. To provide an example of how dance and creative movement might be utilized to transcend language barriers in early elementary classrooms, the San Diego Teaching Artist Project in Southern California will be explored. This project uses dance and theater to boost the English language development of children who speak a language other than English at home. Teacher interviews, observations, and test scores will be utilized to evaluate the success of this strategy.

The impact of demographic change

As the 21st century begins, United States is facing a rate of demographic change that it has not faced since the early years of the 20th century. According to U.S. Census Bureau statistics, in 1970 immigrants made up less than 5 percent of the U.S. population; in 2008 their share of the total had passed 12 percent. Between 2000 and 2010, immigration accounted for roughly one-third of U.S. population growth. Nearly one-fourth of children now have at least one immigrant parent. At the same time, immigration to the United States has shifted away from Europe and toward Latin America and Asia.

Serving this diverse immigrant population presents difficult challenges for American public schools. Between 2000 and 2008, the Hispanic population of the United States grew by 31 percent, with much of the increase coming through immigration; 80 percent of Hispanics lived in the nation's 100 largest metropolitan areas, which also housed 74 percent of the nation's Blacks and 88 percent of Asians. Yet, only 60.8 percent of Hispanics held high school degrees, compared to 80.7 percent of Blacks, 85.1 percent of Asians, and 90.1 percent of Whites.

The impact of these differences in educational attainment is made clear by employment rates. In 2008, employment rates for college-educated Americans were fairly constant across the 100 largest U.S. metropolitan areas, varying between 80 and 90 percent. In contrast, 2008 employment rates for workers with no more than a high school diploma varied widely from city to city, from 55 to 76 percent. When the deep economic downturn of 2009 hit these differences were exacerbated. Between December 2007 and November 2009, the unemployment rate increased by 2.7 percentage points for workers with a four-year degree; by 5.3 points for workers with some college or an associate's degree; by 5.7 points for workers with only a high school degree; by 7.4 points for workers without a high school diploma.

Given projections that European American will represent less than half of the nation's working-age (25 to 64) population in 2050, these widely divergent educational outcomes by race and ethnicity have become a matter

of great national concern. How might public schools effectively educate limited English-proficient students both native born and immigrant? Despite this increased concern, the research literature provides no solid conclusion on how best to educate children who speak a language other than English at home.

Surprisingly, given the many millions of initially non English-speaking children who have acquired literacy in English in the United States, and given the many millions of dollars expended on efforts to evaluate bilingual education programs, straightforward, data-based answers are not available (Snow et al., 1998).

According to the California State Department of Education (CDE) there were 1,553,091 English language learners (ELLs) attending K-12 schools in California in the 2007-2008 school year. ELLs comprised 25% of the California kindergarten population. Yet, despite the increasing number of English learners served, many teachers lack the support, training and experience to effectively teach ELL students. As a result, the likelihood of English learners being able to meet the linguistic and academic criteria required to be reclassified as fluent English proficient – even after 10 years in California schools – is less than 40% (Parrish et al., 2006). This article looks at innovative, movement-based approaches to promoting the English language development of limited English-proficient students during their first three years in California public schools.

Motor behavior and psychological development

The San Diego Teaching Artist Project (funded by a federal grant) sends professional dancers and actors into early elementary classrooms to build cultural bridges through creative movement and drama. These activities, in which children simultaneously hear new words and imitate, have proven successful in boosting English language acquisition as measured by the California English Language Development Test (CELDT). The program is based on research showing the contribution that motor behavior makes to psychological development.

Even as very young infants, children are highly motivated to explore, gain information, attend, and engage their physical and social environments (Gibson, 1987). As Gibson (1988) explains: "We don't simply see; we look." Once young children have learned to crawl and walk, they spend roughly half of their waking hours involved in motor behavior – approximately five to six hours per day (Adolph and Joh, 2007). Perception and motor behavior play a key role in children's experiences and psychological processes (Thelen, 1995).

Bertenthal (1996) has proposed that perception and motor action are interrelated rather than autonomous processes. The interrelation of perception and motor behavior becomes clear when one watches young children in action. On a daily basis, infants who have

learned to walk “take more than 9,000 steps and travel the distance of more than 29 football fields. They travel over nearly a dozen different indoor and outdoor surfaces varying in friction, rigidity and texture. They visit nearly every room in their homes and they engage in balance and locomotion in the context of varied activities” (Adolph and Berger, 2006). As they move from room to room, toddlers are taking in, organizing, and interpreting sensory information.

Throughout their early years, much of learning children experience takes the form of imitation. Long before they learn to speak, infants are predisposed to imitate facial and manual actions, vocalizations and emotionally laden facial expressions (Bard and Russell, 1999). Such imitation involves perception and motor processes (Meltzoff and Moore, 1999), making possible the imitation games in which adult and infant mirror one another’s vocalizations and behavior. Imitation is crucial in the acquisition of cultural knowledge (Rogoff, 1990) and language.

Glancing into a classroom

The San Diego Teaching Artist Project uses processes similar to those by which children first learn language in order to help English language learners (ELLs) acquire a second language during their first years of school. Before discussing the structure of the program and how it was evaluated, it may be helpful to provide a glimpse of a teaching artist working with children. This class took place in a diverse San Diego kindergarten classroom during the fall semester.

Before starting the lesson, the teaching artist shows children how to avoid collisions with classmates by creating their own personal “body bubble.” Each child pretends to blow a soap bubble large enough that he or she can stand inside it with arms outstretched. To keep the bubble from popping, each child must stay far enough away from other children so that their bubbles do not touch. This not only avoids collisions but also enhances each child’s awareness of space.

As the music starts, children mirror the teacher’s motions. Today the focus is on the concepts of comparison and contrast. Awareness of contrast is amplified by reaching high and bending low, or by wiggling and then freezing. Following the movements of the teacher, the children experiment with ascending movements (moving upward like smoke, a flower, a bird) and descending movements (melting, sinking, spiraling). At the same time, children learn the basic vocabulary of dance by carrying out movements that are “high,” “middle,” and “low” with respect to the floor. They also do *axial* movements where the body stays in place (swinging, swaying, wiggling, bending, and stretching).

As they become more comfortable with participation in group movement, the children are invited to experiment

with *locomotor* movements in which they walk, hop, slide, bounce, shuffle, skip, etc. This allows for a greater level of creative spontaneity. By second grade, this locomotor activity will be turned into a grammar lesson by expanding the discussion to the pairing of verbs and adverbs. Any locomotor movement can be modified, at the teacher’s suggestion, by matching it with an appropriate adverb such as *quietly*, *angrily*, *happily*, *smoothly*, *sluggishly* (e.g. *shuffle sluggishly*). But the kindergarteners are focused on vocabulary building.

What the kindergarteners are not aware of is the gradual enhancement of their kinesthetic awareness as they engage in weekly dance activities. Kinesthetic awareness is the ability of the central nervous system to communicate and coordinate parts of our bodies with each other. It encompasses the body’s abilities to coordinate motion as well as the body’s awareness of where it is in time and space. When you see a troupe of ballet dancers moving together it is largely kinesthetic awareness that allows them to move in harmony instead of bumping into each other. In an energetic kindergarten class, this wordless communication builds group cohesion.

The structure of the dance classes provides an opportunity for children to both experience the movements and gain a deeper understanding of the words used to describe the movements. The physicality of dance provides a scaffolding tool to help student build on what they already know. Although ELLs might feel intimidated when called upon while sitting at their desks, such was the atmosphere cultivated in the dance classes that the ELLs readily joined in when asked to perform during a dance lesson, even when they felt uncertain about their abilities.

MATERIALS AND METHODS

The lesson described in the vignette above was developed through a partnership between the University of California, Irvine (UCI) and the San Diego Unified School District (SDUSD). The goal of the project is to provide K-2 teachers with professional development that will enable them to integrate arts instruction with the existing language arts curriculum. Each teacher co-teaches 14 standards-based dance lesson and 14 drama lessons with a teaching artist. The next year teachers implement the same lessons by themselves, without the support of a teaching artist. Teachers also participate in professional development workshops that assist them in integrating these arts-based lessons with the English language arts curriculum.

This article focuses on the dance lessons because the powerful impact of these lessons, especially for kindergartners who entered school with little or no knowledge of English, was a revelation to many participants. Fifteen elementary schools serving high-poverty neighborhoods in San Diego will take part in the project. This article focuses on the five schools that began the project in 2010-11. The lead teachers at each grade level at each school (15 teachers) were interviewed about the effect of this experience on their pupils. The interviews were transcribed and analyzed, looking for themes. Classroom observations were made at participating schools. Student scores on the California English Language Development

Test (CELDT) were compared to a matched control group. The CELDT is a state-mandated, standardized test given annually, at the start of the school year, to English language learners in California. Test scores in the experimental group (267 students; 89.5 percent Latino) and control group (2981 students; 86.2 percent Latino) were analyzed utilizing a gain scores analysis approach. To insure fidelity to the treatment, all teaching artists implemented the same set of lesson plans. In addition, streaming videos (Kindergarten Dance Lessons: <http://www.class.uci.edu/dance.html>; Grade 1 Dance Lessons: <http://www.class.uci.edu/dancefg.html>; Grade 2 Dance Lessons: <http://www.class.uci.edu/dancesg.html>) of an expert teaching each of the first 9 lessons in dance and drama at each grade level were made available on-line so that teachers could refresh their memories (Please see on-line lesson plans and classroom videos at the end of article).

RESULTS

Four broad themes emerged from the analysis of teacher interviews: 1) Before receiving the TAP professional development, most teachers rarely taught standards-based arts lessons and/or did not feel comfortable implementing the arts in their classroom; 2) Teachers indicated that the opportunity to gradually increase their participation in co-teaching arts lessons with a teaching artist contributed to their content knowledge, confidence, and skill in teaching ELLs; 3) Teachers reported that children enjoyed the lessons and that they contributed to group cohesion; shy children often “blossomed” during dance lessons, showing another side of themselves; 4) Teachers reported benefits specific to the English language acquisition of ELLs.

Social-emotional development

A typical teacher’s comment about the dance lessons was: “There are some techniques in dance that were really good and I use them in all areas. I use them to settle the kids down or to quiet them or to refocus them.” A dance strategy that K-2 teachers universally found useful was the personal “body bubble.” This activity, in which children imagined themselves surrounded by a delicate soap bubble they must not pop, helped them become more aware of their bodies in space—and in relation to others around them. The sense of personal space developed during dance lessons also transferred to activities such as walking in a line and sitting on the carpet during group time. Another useful strategy was derived from the Freeze Dance. When the class got a little over-excited, the teacher could just say “Freeze!” and all action stopped.

Another observation heard over and over in the interviews was “the shy children really had a chance to use their talents or skills.” Passive and introverted children found the courage to express themselves during the dance lessons—and often afterwards. One first grade teacher described in considerable detail the effect of the dance lessons on one struggling reader who had been

retained two years in kindergarten:

He was truly talented and felt comfortable during the acting and dance... He would get so into it, he'd kind of stop and look around and would start laughing. That was something I had never seen from him. I do feel like it made him more part of the group... I do feel like students saw his personality in dance and theatre, which they didn't see at other times because reading creates stress for him.

However, some teachers pointed to the calming effect dance had on more aggressive children:

My challenge was with the boys I had this year. They don't like being told what to do by other people. But, with drama and dance, you have to listen to each other and cooperate. Otherwise it doesn't work. The last day with dance, I was really impressed with the boys. They had to come up with a folk dance. They really had to listen to each other and cooperate. I was impressed.

Enhanced students’ engagement

Many teachers commented that their pupils looked forward to the weekly dance lessons and stayed focused during the activity, often mastering skills beyond the level of expectation. Teachers’ observations included:

1. They were super fired up for dance. “Ooh, dance! Is it dance today?”
2. I asked them this morning: “All right, guys, I want to know which one you preferred (dance or drama).” And they said “Dance!” almost overwhelmingly.
3. I put a writing prompt on the board: “What did you like better: dance class or theatre class? Why? Give details.” Eighteen students put dance. Only five put theatre.

These observations were in line with the findings of a study of the pilot project upon which the current version of the Teaching Artist Project is based. This study measured the impact of a similar arts program (which included visual art) on student engagement by comparing attendance on days when arts lessons were offered to attendance on days with no arts lessons. Across the five schools and three grades (K-2) involved at that time, the artists were present in the schools for 293 days; whereas they were not in the schools for 1,727 days. Across the days with an artist, the average attendance rate was 94.41 percent. Across the days without an artist, the average attendance rate was 93.49 percent. Therefore, a first estimate of the effect of the artists on attendance was the difference in these rates, 0.92 of a percentage point. To provide the best estimate of program effect, a regression was run that controlled for school, grade, and month. We found that, on average, student attendance was a statistically significant 0.65 percentage points

higher on days and in locations where the teaching artists were present. When we consider that in these schools and grades, attendance on days with no artist already averaged 93.49 percent (a 6.51 percent absence rate), the increase of .65 meant a 10% reduction in absences on days when the teaching artist was present.

Impact on English language learners

A theme that came out clearly in the teacher's interviews was that, while English learners do face many additional challenges, it is pivotal that teachers strive to teach them the same cognitive skills that are typically taught to English-proficient students. Unfortunately, there was widespread recognition among teachers that their efforts had not always met with. The dance lessons were valued because they enabled ELLs to learn by doing. A teacher observed:

I felt like that was very powerful. My students have a set of vocabulary words that they wouldn't have had otherwise. They are able to use it and understand. For me that was pretty cool. They have this whole set of very appropriate vocabulary words that I hadn't given them access to before this year--complex words that they've absolutely got under their belts. I saw that as a huge change.

Dance activities both provided a non-threatening environment that invited students to explore and also allowed them a means to show what they knew that was not dependent upon English vocabulary. Teachers also reported that the cohort of students who had participated in the dance lessons reached more advanced stages of language development earlier in the school year than previous cohorts. The teachers all had examples of how they integrated dance. In the following quote, a second grade teacher described the benefits her class derived from dance:

The dance terms really transferred into our curriculum, like 'high,' 'medium,' and 'low,' also 'fast,' 'slow,' 'heavy,' and 'light.' Those are also all math terms that they act out in a kinesthetic way. So, when we actually get to terms like 'heavy' and 'light' in math, they already know it and understand it because they used their bodies to understand the terms rather than just hearing them.

A first grade teacher commented:

Dance fit in really well with what we were doing this time of year. We were studying animals, and so it lent itself to talking about what kind of energy and movement an animal would have.

A kindergarten teacher explained:

After dance, the discussion was, "Can we move from the rug to line up for recess in tiptoe? Can you do it twirling?"

Measuring English Language development

The teachers interviewed felt that the effect of the program, in terms of English language development, was significant. But how accurate were their perceptions? To determine the impact of the K-2 Teaching Artist Project on student listening and speaking skills, a quasi-experimental study was designed. The California English Language Development Test (CELDT) was used to measure increases in student language abilities. When the study began, the participating schools had been randomly selected from 40 diverse schools. This section describes the analysis of Year 1 data from five experimental schools and 13 randomly selected comparison schools.

CELDT data for the fall 2010 and fall 2011 administration were provided by the San Diego Unified School District (SDUSD), along with demographic data that could be linked to each student participant. These files represent students who could be matched across two CELDT scores from fall 2010 to fall 2011 representing students in first, second, and third grades for the fall 2011 CELDT administration. The study included an experimental group consisting of 267 students and a comparison group consisting of 2981 students.

ANOVA was utilized to look at baseline differences between comparison and experimental students on CELDT listening and speaking scores for fall 2010. The only significant difference was ethnicity, but it should be noted that both samples of comparison and experimental students are more than 85 percent Hispanic. The difference lies in the 15 percent of students in the comparison and experimental groups with the comparison group having a significantly greater number of Asian, Pacific Islander, and Filipino students than the experimental group.

The analysis of the CELDT fall 2010 and fall 2011 listening and speaking score analysis utilized a gain scores analysis approach. Using this approach the fall 2010 listening and speaking scale scores were subtracted from the fall 2011 listening and speaking scale scores to produce a gain score between the two years. The gain score was then converted to a Z-score (Rogosa and Willett, 1983), which is a commonly used approach for multi-year analysis using an ANCOVA framework. The analysis showed a significant positive difference in both listening and speaking for the experimental group when controlling for the 2010 scaled scores.

DISCUSSION

In the United States, nearly 70 percent of limited English

proficient (LEP) students are enrolled in only 10 percent of elementary schools (Cohen, 2005). In these predominantly urban schools, LEP students account for almost half of the student body (on average). Since these schools tend to be located in neighborhoods where most residents speak a language other than English at home, these children may have limited opportunity to learn oral English in informal settings. Therefore, opportunities to interact with adults at school and receive feedback are crucial. The dance and drama activities in the Teaching Artist Project allowed teachers to verbally interact with many children at once. Watching their responses, teachers were able to 1) assess a wide range of student learning and 2) quickly and informally correct misunderstandings.

Throughout history, dance has been used to educate, celebrate, and symbolize (Hanna, 2008). Dance helps to enhance our understanding of the power of a part of the human toolkit that we too often ignore. For young children, both ELLs and native English speakers, dance provided a means of transcending the limitations of their linguistic knowledge. Teachers involved in the Teaching Artist Project repeatedly commented that the dance lessons had revealed aspects of the personalities of the children in their classes that they had never glimpsed before. Although the focus of this article has been on addressing the challenges faced by ELLs, increased awareness of the signals we continually send to one another through movement, gesture, posture, and facial expression could provide clear advantages for teachers.

During initial implementation of the project, many teachers showed a strong preference for the drama activities. As one teacher pointed out, "Theatre felt more natural. Dance, I'm still not comfortable with. Dance language and movement are still not intuitive to me." This was because the drama activities were closely aligned with the reading curriculum. Acting out scenes from stories as a way to help children better understand plot and character made intuitive sense to teachers; whereas, teachers associated dance with the folk dances taught in physical education classes. However, over time, the children's enthusiasm for the dance lessons won the teachers over. The skills learned in the theater lessons may have transferred more easily to classroom literacy activities; they did not generate the same level of engagement.

Conclusion

Non-verbal communication has long been known to promote human development, learning and knowledge (Corballis, 2002). Greenland (2000) points out that movement conversations are the first conversations we have. If you watch a baby in its mother's arms, you will see the mother communicates her approval or her adoration of the baby, by mirroring the tiny movements the baby makes. She will communicate disapproval through

movements that oppose the baby's, that show difference. She tunes into and communicates with the baby through movement. Mother and baby do this unconsciously; it is part of the human package. We continue to use the same methods of tuning in, communicating agreement and disagreement, all our lives. We have movement conversations with one another, even though we may choose to ignore them. For teachers, the challenge is learning to see and act upon movement signals.

In *Origins of the Modern Mind*, Donald (1991) outlined stages of human cultural development that he called Mimetic, Mythic, and Theoretic. He argued that, before the invention of language, human beings possessed imitative capacities that enabled them to collectively invent and maintain customs and skills; therefore, each member of the social group did not have to reinvent the knowledge and skills on which survival depended. This way of looking at human development differs markedly from the fashionable view that human understanding is essentially language understanding, that we are "incarnated vocabularies" (Rorty, 1989). Still, it is a view for which the research cited earlier on the role of motor behavior in development – along with casual observations of infant development – provides support.

There is much evidence to suggest that dance taps into something beyond language, a kinesthetic relationship to the world that is foundational to all later understanding. All children may therefore benefit from integration of creative movement into the school literacy curriculum. However, for children entering a classroom where the language they speak at home is not understood, the opportunity to fall back on an alternative language of movement has special importance. For them, dance can open a doorway to understanding.

On-Line Resources for Teaching Artist Project

Kindergarten	Dance	Lessons:
http://www.class.uci.edu/dance.html		
Grade 1	Dance	Lessons:
http://www.class.uci.edu/dancefg.html		
Grade 2	Dance	Lessons:
http://www.class.uci.edu/dancesg.html		

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

- Adolph KE, Berger SE (2006). Motor Development. In Handbook of Child Psychology: Volume 2: Cognition, Perception, and Language (Sixth edition). Series Editors: W. Damon and R. Lerner. Volume Editors: D. Kuhn and others. New York: John Wiley and Sons.
- Adolph KE, Joh AS (2007). Motor development: How infants get into the act. In Introduction to Infant Development (Second edition). Edited by A. Slater and M. Lewis. New York: Oxford University Press.

- Bard K, Russell C (1999). Evolutionary Foundations of Imitation: Social-Cognitive and Developmental Aspects of Imitative Processes in Non-Human Primates. In *Imitation in Infancy: Cambridge Studies in Cognitive and Perceptual Development*. Edited by J. Nadel and G. Butterworth. Cambridge, UK: Cambridge University Press.
- Bertenthal BI (1996). Origins and Early Development of Perception, Action and Representation, *Annu. Rev. Psychol.* 47:431-459.
- California Department of Education (2008). DataQuest. Retrieved February 22, 2009, from <http://data1.cde.ca.gov/dataquest/DQReports.asp?CDSType=S&CDSCode=38684783830288&IPage=P>
- Cohen C (2005). Who's left behind? Immigrant children in high- and low LEP schools. Washington, DC: Urban Institute.
- Corballis MC (2002). *From hand to mouth: The origins of language*. Princeton, NJ: Princeton University Press.
- Donald M (1991). *Origins of the modern mind*. Cambridge, MA: Harvard University Press.
- Gibson EJ (1987). What Does Infant Perception Tell Us About Theories of Perception? *J. Experiment. Psychol: Human Perception and Performance* 13(4):515-523.
- Gibson EJ (1988). Exploratory Behavior in the Development of Perceiving, Acting and the Acquiring of Knowledge, *Annu. Rev. Psychol.* 39(1):1-41.
- Greenland P (2000). *Hopping home backwards: Body intelligence and movement play*. Leeds, UK: JABADAO Centre for Movement Studies.
- Hanna JL (2008). A nonverbal language for imagining and learning: Dance education in K-12 curriculum. *Educ. Res.* 37(8):491-506.
- Meltzoff AN, Moore MK (1999). Persons and Representation: Why Infant Imitation Is Important for Theories of Human Development. In *Imitation in Infancy: Cambridge Studies in Cognitive and Perceptual Development*. Edited by Nadel J, Butterworth G. New York: Cambridge University Press.
- Parrish TB, Perez M, Merickel A, Linqanti R (2006). Effects of the Implementation of Proposition 227 on the Education of English Learners, K-12: Findings from a Five-Year Evaluation. Sacramento, CA: American Institutes for Research and WestEd.
- Rogoff B (1990). *Apprenticeship in Thinking: Cognitive Development in Social Context*. New York: Oxford University Press.
- Rorty R (1989). *Contingency, irony, and solidarity*. Cambridge: Cambridge University Press.
- Snow CE, Burns MS, Griffin P (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Thelen E (1995) Motor development: A new synthesis. *Am. Psychol.* 50(2):79-95.

Full Length Research Paper

Reliability and validity of test of gross motor development-2 (Ulrich, 2000) among 3-10 aged children of Tehran City

Farrokhi A.¹, Zareh Zadeh M.², Karimi Alvar L.³, Kazemnejad A.⁴, and Ilbeigi S.^{3*}

¹Faculty of Physical Education, University of Tehran, Tehran, Iran.

²Faculty of Physical Education, University of Kerman, Kerman, Iran.

³Faculty of Physical Education, University of Birjand, Birjand, Iran.

⁴Faculty of Physical Education, University of Tarbiyat Modares, Tehran, Iran.

Received 24th January 2012; Accepted 24th January 2013; Published May 2014

The purpose of this study is to determine construct validity and three reliability aspects of Test of Gross Motor Development (TGMD-2; Ulrich, 2000) in Tehranian children aged 3:0-10:11. TGMD-2 which includes two subtests (locomotor and object control) is designed to assess movement pattern development of 12 fundamental movement skills. The TGMD-2 was administered to 1438 children. Internal consistency reliability for locomotor and object control score and also for total composite score averaged 0.78, 0.74 and 0.80, respectively. Internal consistency reliability was high for females and males, too. Test-retest reliability ranged from 0.65 to 0.81, and intra-rater reliability was above 0.95. To investigate construct validity, Ulrich's (2000) two-factor model was postulated and this hypothesis was tested through confirmatory factor analysis. According to the model, six variables or skills measuring child's ability for moving into space were loaded on one factor (locomotor), while the other six variables measuring ability for controlling and manipulating objects were loaded on the other factor (object control). According to the findings, two-factor structure of TGMD-2 and also proper assignment of skills to locomotor and object control factors were supported for our population, too. Additionally, the correlation coefficients between age and subtests' score provided support for another aspect of construct validity, that is, developmental nature of TGMD-2; the resulting coefficients indicated that TGMD-2 is capable of differentiating between ages. In conclusion, this study indicated that TGMD-2 can be used with confidence to assess gross motor development of the studied population.

Key words: TGMD-2, construct validity, internal consistency, test-retest, intra-rater reliability.

INTRODUCTION

The principal element in motor development is the fundamental movement skills (Reeves et al., 1999), which are

included in the gross movements, that is, the movements which are related to the large force producing muscles

*Corresponding author. E-mail: saeed.ilb@gmail.com. Tel: +985612502124. Fax: +985612502124.

Author(s) agree that this article remain permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

function (Thomas and Nelson, 1996). These skills which refer to the purposeful movement patterns include locomotor and object control movements. The locomotor movement patterns include the skills used for the purpose of transporting the body through space, like running and jumping. The object control movement patterns, however, are the skills that involve giving force to object or receiving force from objects, like throwing and catching the ball. These skills develop, in early childhood, as a function of physical maturation and practice and are the prerequisites for games and specialized movements (Gallahue and Ozmun, 2006).

The mastery of certain fundamental movement skills is a prerequisite for daily life functioning and participation in later physical or sport-specific activities. At an early age, gross movement skills are necessary to move, stabilize and control body and objects while exploring the environment. Later in life, well developed gross movement skills help individuals to function more smoothly (Cools et al., 2008). Studies have indicated that inadequate competence in these skill causes lack of success and the feeling of being incapable in games (Vira and Ruadsepp, 2003) which leads to the gradual elimination of physical activity through the life span (Deflandre et al., 2001) and the improper level of physical fitness (Reeves et al., 1999). In addition, children who are weak in these skills cannot easily establish social relations with others and have a higher level of anxiety and lower levels of self-esteem in comparison to more skillful peers (Piek et al., 2006). Also, delay in gross motor development usually is followed by visual perception disorders (Skordilis et al., 2004), dyslexia and linguistic disorder which, if not diagnosed on time, lead to learning and cognitive problems (Merriman et al., 1995). Thus, one aspect of the screening programs in early childhood should be measuring the development of fundamental movements which calls for the existence of standardized tests in this area. These tests can be divided into two categories, each with advantages and disadvantages, the measurement of motor performance quantitatively by speed, distance or number of successful attempts and the measurement of fundamental movement skills qualitatively by criteria for passing or technique components of the movement. In recent years, the most frequently used fundamental movement skills assessment tools with children employ qualitative measures, because the major advantages of qualitative assessment are the information can be used to inform the teacher which specific components of a skill an individual needs to practice, and the assessment can be undertaken in a more meaningful context than quantitative methods (Hands, 2002). In addition, technique components are not fully acquired by preschool, and their performance is easily influenced by testing conditions in quantitative methods. So it is not always valid to use quantitative scales with preschool children. In tests based on pass or fail criteria, however, motor ability is estimated

by a specific performance criterion, movement patterns can be observed in daily life and special measurement tools and conditions are not needed (Goshi et al., 1999).

One of the best known tests in movement assessment area is Test of Gross Motor Development-2 (Ulrich, 2000) designed to measure the development of fundamental movements based on qualitative aspects of the skills in children of 3 to 11, and includes six locomotion skills and six object control skills. According to the author, the test can be used to identify children who are significantly behind their peers in gross motor performance, to plan programs to improve skills in those children showing delays and to assess changes as a function of increasing age, experience, instruction or intervention (Cools et al., 2008).

One of the qualities of a standardized test is validity. A valid test is one that theoretical and empirical evidences confirm the usefulness and appropriateness of interpretations and applications based on the test results (Van Waelvelde et al., 2004). One aspect of validity is construct validity. Construct is the characteristic reflected in the test scores (Burton and Miller, 1998), like intelligence or motor development. Construct validity refers to the extent to which the underlying construct (factors) or trait, claimed to be measured by the test, can be identified (Thomas and Nelson, 1996) and also refers to the extent to which test scores reflect the construct-related theoretical concept on which the test is based (Saif, 2004). To examine the construct validity, generally hypotheses are generated about the defined constructs assumed to be latent in test performance, and the hypotheses are verified by logical or empirical methods. To identify the number of constructs and the structural model on which the test is based, confirmatory and exploratory factor analyses are used (Ulrich, 2000). In the exploratory factor analyses, using the covariance analyses in the matrix of correlation between the test's measures, the measures that their variations are correlated are grouped as one factor. So, the underlying constructs are extracted and the factorial structure of the test is identified (Sarmad et al., 2004). But, the identified factorial structure is specific to the population the test is designed for. Therefore, the appropriation of the test's factorial structure in a new population is evaluated through confirmatory factor analysis and goodness of fit indexes (Thomas and Nelson, 1996).

The Test of Gross Motor Development was originally developed in the United States for typically developing (TD) children but it has been translated and validated in different countries for children with and without disability. Evidence of the construct validity of the TGMD-2 was reported in its test manual. TGMD-2 was validated on 1,208 American children via exploratory factor analysis using principal component analysis with promax rotation and confirmatory factor analysis with maximum likelihood mode of estimation. Although the exploratory factor

analysis results identified two factors of locomotor skills and object control skills, the loadings of two items (Strike and Jump) were unclear. The fit indexes provided by confirmatory factor analysis indicated that the two-factor model of the TGMD-2 produced a good approximation to data; however, the model did not represent a reasonable fit ($\chi^2 = 280.3$, $df = 53$, $\chi^2/df = 5.29$). In addition, the level of significance and some fit indexes such as root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), non-normed fit index (NNFI) and comparative-fit index (CFI), as well as, the path diagram with item loadings and correlation between two factors, provided a better understanding toward the underlying structure of a testing instrument, was not mentioned in the manual (Spessato et al., 2012). Wong (2006) tested 1251 TD Hong Kong Chinese children. A principal axis factor analysis with oblique rotation determined two factors, but five items were loaded on both factors. After eliminating these items progressively and computing additional principal axis factor analyses, the desirable exploratory factor analysis results were two factors with eight items loaded satisfactory on factors suggested by Ulrich (2000). Four two-factor models with 12-, 11-, 9-, and 8-item were tested using confirmatory factor analysis. Although the two-factor model of the TGMD-2 did not represent exact fit, goodness-of-fit indexes supported that the four models were tenable (Wong, 2006). For Brazilian Children Appropriate indices of the confirmatory factorial validity and adequate items correlations with the correspondent subtest were found, that provided reasonable support for the TGMD-2 two factor model (Valentini, 2012).

TGMD-2 has also been used to test children with sensory disability and those cognitively impaired. Satisfactory fit indexes were reported for Flemish children with cognitive disabilities (goodness-of-fit index [GFI] = .88; adjusted goodness-of-fit index [AGFI] = .82) (30) and visual impairment children from Netherlands (RMSEA = 0.07, GFI = 0.85) (Houwen et al., 2010).

Extensive research addressed the influence of age and gender on the proficiency of gross motor skills. It was reported that gross motor skills improved with age (Wong, 2006). So, In addition to factorial structure analysis, the other aspect of construct validity of the tests having construct with developmental nature, such as TGMD-2, is the investigation of age differentiation capability or the score improvement by age increasing (Saif, 2004). Thus, it could be hypothesized that there is a high correlation between the children's test performance and age. One way of establishing a test's construct validity is to study the performances of different groups of individuals on the test (Ulrich, 2000). Apart from age related differences, gender differences in gross motor skill performance have been established in both children and adolescents, with boys achieving higher scores than girls especially in object control skills (Okely and Booth,

2004; Spessato et al., 2012). Then with reference to the literature, it would be expected that boys would obtain higher object control skill scores. Ulrich (2000) indicated that both subtests were strongly related to the chronological age ($r=0.69$ to 0.72 , $p<.05$ for locomotor; $r = 0.71$ to 0.75 , $p<.05$ for object control). Ulrich did not report gender differences in the manual but the reported raw score means showed boys did better than the girls on the object control subtest (Ulrich, 2000). Niemeijer (2007) also, observed high correlations between age and subtests total score, ranged .66 to .81, among typically development Dutch children (Niemeijer, 2007). Simons et al. (2007) reported significant but low correlation between the age and object control subtest scores ($r = 0.25$) for children with mental retardation; ANOVA also revealed a significant age effect in the object control subtest but not in the locomotor subtest. Significant effect of gender was also observed in object control skills among mental retarded children (Simons et al., 2007). For children with visual impairment, there were significant partial correlations between age and the locomotor ($r = 0.36$, $p = 0.002$) and object control ($r = 0.53$, $p<.001$) subtests. A significant effect for gender in the object control subtest was found, but not in the locomotor subtest (Hands, 2002).

Reliability is the other feature of standardized tests and the prerequisite for validity, that is, an assessment instrument that is not reliable cannot be valid (Saif, 2004). Reliability pertains to repeatability and consistency of test scores. Theoretically, an observed score (X) is compromised of an examinee's true score (T) and error score (E), that is, $X = T + E$. Reliability, then, is the degree to which observed test scores match true score. With regard to psychometric definition, reliability is presented in terms of the variance components (S_T^2 / S_X^2). As the error variance decreases, greater proportion of the observed-score variance accounted for by the true-score variance, and reliability will increase (Burton and Miller, 1998). Thus, reliability reflects the degree to which test scores are free of measurement error or error variance (Thomas and Nelson, 1996).

Measurement error comes from three sources: content and questions, test administration conditions and scoring. To determine the content reliability of a test, internal consistency is examined. Internal consistency refers to the extent of test's item consistency and congruity (Okely and Booth, 2004). The error of administration is estimated through test-retest method. Particularly when the performance is expected not to change during a short time, the stability of performance is investigated during a given time period by test-retest (Saif, 2004). In this method, the test is administered at two different times but with the same conditions and the correlation between two set of scores is reported as "stability reliability" (Wong, 2006). The inter-rater and intra-rater reliability are two ways of

investigating the scoring reliability or objectivity. In this method, which requires video-recording, the performance is scored by two raters or by one rater at two different times, then the correlation between the scorings is calculated (Burton and Miller, 1998).

The degree of reliability is expressed by a correlation coefficient, ranging from 0.00 to 1.00, and generally coefficient of above .70 is good. The correlation coefficient is calculated through intra-class correlation (R) and Pearson correlation (r). Some rationale was presented for using ICC methods instead of Pearson r, such as ICC uses ANOVA to obtain reliability coefficient and changes in means and standard deviations from trial to trial or item to item can be assessed in the ICC. So it provides precise estimation of error variance. The best-known technique of estimating reliability based on ICC is Cronbach alpha coefficient (Morrow et al., 2005).

Satisfactory reliability results have been reported for locomotor, object control and composite score of TGMD-2 for different populations. For American children, internal consistency alpha coefficient was higher than .85 (locomotor $\alpha = .85$, object control $\alpha = .88$, and GMQ $\alpha = .91$). Pearson correlation coefficient for the test retest was .88 for locomotor, .93 for object control and for total test .96. Inter rater reliability was .98 for all three aspects (Ulrich, 2000). Desirable test-retest values ($r = .83$ to $.91$) and inter- and intra-rater reliability ($\alpha = 0.86$ to 0.99) were found for TD Brazilian children (Valentini, 2012). The internal consistency of the TGMD-2 was found to be high ($\alpha = 0.71$ to 0.72) and the inter rater, intra rater, and test retest reliability were acceptable (ICCs ranging from 0.82 to 0.95) for Netherlands children with visual impairment of Netherlands (Houwen et al., 2010). Wong and Cheung (2006) also reported acceptable indices of internal consistency ($\alpha = .78$ to $.85$) and Inter rater reliability ($r = .82$ to $.94$) for TD Chinese (Wong, 2006). Alpha values ranged from .82 to .90 for subtests and composite score internal consistency among Flemish children with cognitive disabilities. In addition, Spearman correlation coefficients for test retest reliability of locomotor and object control subtests were .90 for and .92, respectively (Simons et al., 2007).

Although there are evidences of reliability and validity of TGMD-2, the evidence of a measure's reliability and validity is sample-dependent and cannot be generalized to other cultural and geographical areas (Yun and Ulrich, 2002). This fact limits the applicability of TGMD-2 about Iranian children. Considering the sociocultural differences observed across different countries, application of a test in a population with different set of characteristics should be followed by the assessment of reliability and validity of the test to guarantee the correctness of the result (Chow et al., 2006). In Iran, there was no test available to measure the qualitative process of gross motor skills and the introduction of the TGMD-2 to kindergartens and primary schools seemed to be desirable. Due to this, and with

regard to the unpleasant consequences of delayed motor development and the necessity of reliable and valid test existence for motor development assessment, the present study aims at determining the construct validity and three aspects of the reliability of TGMD-2: internal consistency, stability and objectivity among Tehranian children of 3 to 11.

METHODOLOGY

Sample selection

The population was 3 to 11 years old children of public kindergarten and elementary schools, supervised by Education Organization of Tehran City. The sample was selected through a stratified cluster sampling procedure. Based on information from the Education Organization, the sample was stratified by geographic region and education districts size as well as age and gender. Finally, the sampling procedure resulted in 1438 persons as target population for normalization and validation. All the children recruited for testing were healthy and free from diagnosed orthopedic, neurologic, physical impairments, developmental conditions and learning disability.

63 children were selected from the sample for the evaluation of test-retest reliability. Intra-rater reliability was evaluated by testing 32 children which were selected just for this purpose.

Instrument

Test of Gross Motor development-2 (Ulrich, 2000) was the instrument used for collecting the data. This test is a revised version of TGMD (Ulrich, 1985). Ulrich standardized and validated this test using a sample of 1208 American children. TGMD-2 is a process oriented test which assesses development of movement pattern of 12 fundamental movement skills. The age range from 3 to 10 years old covers the period in which the most dramatic changes in a child's gross movement skill development occur. The test includes two subtests: object control and locomotor. Locomotor subtest includes running, galloping, hopping, jumping, leaping and sliding while object control subtest includes striking, dribbling, catching, kicking, throwing and rolling.

Each skill is evaluated based on some performance criteria. The content of each criterion is about one component of the advanced pattern of skill performance. Each subtest includes 24 performance criteria. For scoring, the child has to perform the task at two trials. Each criterion is given a 1, if the criterion is performed correctly; and a 0, if performed incorrectly. The test is administered in 15 to 20 min. Each subtest has a total raw score attained by summing the scores for the related skills. The maximum score is 48. The normative scores include the standardized score for the subtests, with mean and standard deviation of 10 and 3 respectively and also the Gross Motor Quotient which is a composite score based on the standardized scores (Ulrich, 2000). GMQ which has a mean and a standard deviation of 100 and 15 respectively is the best measure of a child's overall gross motor ability (Burton and Rodgeron, 2001).

Data collection procedure

The study was approved by research and ethical committee of Tehran Organization of Education, and the investigator got the

permission to perform the TGMD-2 in selected school. The school members were informed about the purpose and safety measures. After admittance of the school, students' date of birth, parents' education and health status were obtained from personal files and the information was considered in sampling procedure. Prior to data collection, the children's parents gave informed consent for their children's participation.

To assure measurement consistency, the same tester (investigator) observed and scored all children's performance. Prior to testing, training was done on administrating and scoring of the TGMD-2 based on the manual instructions. Before the actual data collection, a pilot-testing was conducted in order to evaluate the investigator's scoring competency and intra-rater reliability. For estimating intra-rater reliability, the investigator watched and scored the videotaped performance of selected participants twice, with a 12-day interval. The performance was video recorded by using Sony R, CCD-TRV418E, 1322101 Camcorder. The TGMD-2 was administered twice, with a 2-week interval between testing, for test-retest reliability. The testing was done during the physical education class. Arrangements were made to accommodate the test in a safe environment for the children in order to minimize administration time and distractions. The children were tested in small groups and each child finished the test in one session. The testing followed standardized test procedures as provided in the test manual (Ulrich, 2000).

Data analysis

The reliability of TGMD-2 was investigated through internal consistency, test-retest and intra-rater reliability methods. The scores of the entire normative sample served as subjects for internal consistency analysis and the data for evaluating stability and objectivity reliability were obtained from the sample selected for these two purposes. Alpha coefficient was computed for internal consistency on the subtest and GMQ using SPSS. But, internal consistency for GMQ was calculated using the following formula which is designed for composite scores (20).

$$\alpha = 1 - \frac{\sum \delta_{xi}^2 (1 - \rho_{xixi})}{\delta_z^2}$$

δ_{xi}^2 = subtests score's variance, ρ_{xixi} = subtest's reliability coefficient, δ_z^2 = composite score's variance.

Internal consistency coefficient was computed for age groups separately and in order to estimate total internal consistency reliability of subtests and GMQ regardless of age, the alphas were averaged using z-transformation method, the formula for calculating the mean of some correlation coefficients (Thomas and Nelson, 1996). Intrarater and test-retest reliability was evaluated using the intraclass correlation coefficient (ICC). To control any effects of age in the selected samples on the evaluation of test-retest and intra-rater reliability, the total score of each subtest was changed into standardized scores.

The construct validity of TGMD-2 was investigated through confirmatory factor analysis and examining age differentiation and gender differences. Maximum-likelihood confirmatory factor analyses were performed to test the goodness-of-fit of TGMD-2 skill assignment to the subtests, using Lisrel (Lisrel 8.8, Scientific Software International, Lincolnwood, IL, USA). It was assumed that

Table 1. Alpha coefficients for TGMD-2 scores at age intervals and gender.

Age (N)	Subtests		GMQ
	Loc	OC	
3 (159)	.88	.72	.86
4 (178)	.87	.69	.83
5 (190)	.83	.69	.80
6 (190)	.77	.75	.81
7 (184)	.75	.76	.77
8 (177)	.66	.77	.79
9 (187)	.74	.78	.84
10 (173)	.65	.71	.72
Avg. α	.78	.74	.80
Girl (719)	.92	.85	.91
Boy (719)	.91	.89	.92

the locomotor subtest includes the six items of running, galloping, hopping, jumping, leaping and sliding and object control subtest includes the six items of striking, dribbling, catching, kicking, throwing and rolling. Pearson correlation coefficients were calculated to examine the relationship between age and subtests total score. A 4x2 (Age x Gender) two-way analysis of variance was performed to examine further age related developmental changes and also gender differences on subtest performance. For the analysis, age bands were defined as age band 1, from 3-4 years; age band 2, from 5-6 years; age band 3, from 7-8 years; age band 4, from 9-10 years.

RESULTS

TGMD-2 reliability

Table 1 shows the internal consistency reliability coefficients for the locomotor subtest (Loc), object control subtest (OC) and GMQ, and also shows the averaged coefficients. At the end of the table alpha for the two groups of boys and girls has been provided.

As shown in Table 1, the range of the internal consistency reliability coefficients for the eight age groups is from .65 to .88. The averaged alpha coefficients show that the total reliability is above .74. In addition, the alpha for each of the genders is above .85. Although the acceptable value of reliability coefficient is above .70 (4), the minimum acceptable level for alpha has been reported to be .65 (1). Thus, the resulted alpha values confirm internal consistency reliability of TGMD-2 for Tehranian children of 3 to 11.

Table 2 shows the ICCs and corresponding 95% Confidence intervals for, intrarater and test-retest reliability. In this table, the mean and standard deviation of the subtests' total score for each testing and scorings are provided. The content of Table 2 shows a little difference

Table 2. The test-retest and intra-rater reliability coefficient of TGMD-2.

TGMD-2 scores	Test-retest			Intra-rater		
	TM (SD)	TM (SD)	ICC(95%CI)	TM (SD)	TM (SD)	ICC(95%CI)
Loc	32 (3)	33 (3)	.65(.50-.79)	32 (3)	31(3)	.95(.91-.97)
OC	20 (6)	21 (5)	.85(.75-.91)	24 (5)	24 (4)	.99(.97-.99)
GMQ	100 (15)	100 (15)	.81(.70-.89)	100 (15)	100 (15)	.97(.94-.98)

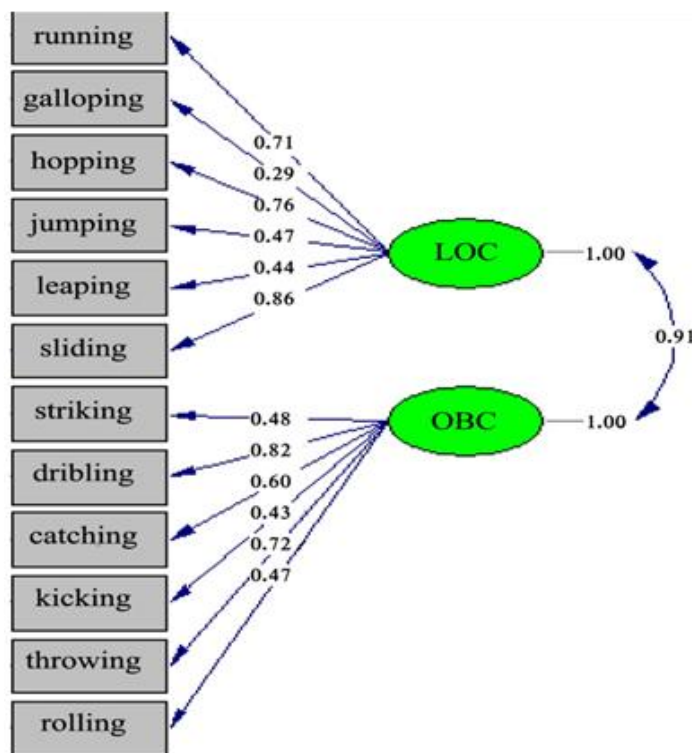


Figure 1. Pictorial profile of hypothesized two-factor model of TGMD-2.

between the means for each testing and scorings. for interpretation of the resulted coefficients we adopted the criteria recommended by Fleiss (1981) that consider ICCs of > 0.74 as excellent, < 0.40 as poor, 0.40-0.59 as fair and 0.60-0.74 as good. Based on the criteria, the coefficients for test retest and intrarater reliability were at good to excellent level (Fleiss, 1981).

Validity of TGMD-2

Validity was evaluated using confirmatory factor analysis. Figure 1 shows the schematic representation of the two-factor structure of the test and the internal relations of the components.

According to Figure 1, the loading of 12 skills on two factors (locomotion and object control) range from .29 to

.86. The least value belongs to galloping and the highest value belongs to sliding. As the loading values below .30 are considered low (11), the resulted values shown in Figure 1 are desirable. All factors loadings were statistically significant ($p < .05$). The figure shows that the correlation between the two factors is high which indicates that a one factor solution would be appropriate for the data rather than a two factor solution. Thus, to investigate the correctness of the assumed factor structure, goodness of fit indexes are taken into account which is shown in Table 3.

The overall fit of the data to the evaluation of the proposed factor structure is initially based on the non-significance of chi-square (χ^2) (Evaggelinou et al., 2002). A non-significant p value for χ^2 means that the model is a good representation of the data and there is no reliable difference between the sample matrix (that is, the data)

Table 3. The values of goodness of fit indexes for the two-factor structure of TGMD-2.

Model	Fit Indexes								
	χ^2	df	χ^2/df	GFI	AGFI	RMSEA (90% CI ^{**})	NNFI	CFI	SRMR
Two-factor model	303.9 [*]	53	5.70	.97	.95	.05 (.05-.06)	.97	.98	.03

*p<.05 **Confidence Interval.

Table 4. Subtests' total score means (and standard deviations) and age differentiation correlation coefficients.

Age	Loc		OC	
	Girl	Boy	Girl	Boy
	M(SD)	M(SD)	M(SD)	M(SD)
3	16(6.5)	17(6.5)	12(3)	13.5(4)
4	24.5(6)	25.5(6)	15(3.5)	17(4)
5	28.5(5)	29(5)	17(4)	19(4)
6	31.5(4)	31(3.5)	18.5(4)	22(5)
7	32(3.5)	33(3.5)	21(4.5)	26.5(5)
8	33.5(3)	33(3)	24(5)	28(5)
9	34(3)	34(4)	25(5)	30(5.5)
10	34(3.5)	34.5(2.5)	27(4)	32.5(4)
Correlation with age (r)	.68 [*]	.66 [*]	.76 [*]	.79 [*]

*p<.05.

and the fitted matrix (that is, the model) (Houwen et al., 2010). As shown in Table 3, this index is significant here. However, χ^2 is strongly affected by sample size and in large samples there is likelihood that it would be significant (Evaggelinou et al., 2002). Thus, more valid indexes are used besides χ^2 (Cools et al., 2008). The values of 2 to 5 for χ^2/df show the reasonable fit of the assumed factorial model (Ulrich, 2000). The goodness-of-fit index (GFI) and adjusted goodness-of-fit index (AGFI) were examined to provide information about the proposed model estimate covariance and sample covariance (Valentini, 2012). The GFI was considered to be a measure of the proportion of variances and covariances that the proposed model was able to explain (Wong, 2006). Both GFI and AGFI indexes range from 0 to 1 and values greater than .90 are indicative of a good model (Evaggelinou et al., 2002). In terms of assessing the degree of lack of fit of the model, the root mean square error of measurement (RMSEA) was computed. For the examined model, the values between .05 and .08 indicate a reasonable approximation to the data (Wong, 2006). RMSEA values less than .05 would indicate a close fit of the model while values of about .08 would indicate a reasonable fit (Evaggelinou et al., 2002). The comparative-fit index (CFI) was used to estimate model discrepancies. The CFI compared the proposed model with the null model assuming no relationships between measures.

The non-normed fit index (NNFI) was computed to examine the proportionate improvement in fit of the model compared to a baseline model in which all observed variables are uncorrelated. The values of NNFI and CFI in the .90 or above were considered as a reasonably good approximation of the data. The standardized root mean square residual (SRMR) was computed for the proposed model to provide a summary measure of standardized residuals. The small values of the SRMR for the model (<.05) demonstrated an acceptable fit (Valentini, 2012).

As shown in Table 3, although χ^2 index is significant, the fit indexes are quite acceptable with respect to the criteria reported above. GFI and AGFI considered as absolute fit indexes (Evaggelinou et al., 2002) are close to 1, reflecting reasonably good approximations of the data. RMSEA value, another absolute fit index, demonstrates that the two factor structure of TGMD-2 was close fitted. The values of other indexes included as descriptive-fit and alternative-fit indexes provide additional support for the fit of the two-factor model.

Table 4 shows the correlations between age and the subtests' scores of TGMD-2. The means and the standard deviations for each subtest are presented. Because of the developmental nature of motor development, it was expected that there is a high and significant correlation between age and the total score of the two subtests.

Table 4 shows that the mean of the performance of the sample in the subtests improves as the age goes up. The obtained coefficients also show the TGMD-2 subtests are strongly related to age.

Factorial ANOVA revealed a significant main effect for gender in the object control subtest, $F(1, 1430) = 228.06$, $p < 0.001$, but not in the locomotor subtest, $F(1, 1430) = 2.79$, $p = 0.09$, with the boys doing significantly higher than the girls on the object control subtest. The results showed significant influence of age for locomotor subtest, $F(3, 1430) = 465.27$, $p < 0.001$, and object control subtest, $F(3, 1430) = 648.52$, $p < 0.001$. Post hoc analysis of influence of age using Tukey HSD test revealed that significant differences ($p < 0.001$) between the four age bands (3.0 to 4.0, 5.0 to 6.0, 7.0 to 8.0, and 9.0 to 10.0), with the elder bands outperformed younger bands on the performance of both subtest.

DISCUSSION

TGMD-2 reliability: In investigation of the internal consistency reliability of TGMD-2, the alpha coefficient values for the two subtests of locomotion, object control and the composite score of GMQ are .78, .74 and .80 respectively. As the values of higher than .70 are interpreted as good and those above .80 are interpreted as very good (Armitage and Theodore, 1998), the findings are indicative of the fact that the internal consistency reliability of TGMD-2 is acceptable for 3 to 11 years old children of Tehran City. It reflects the homogeneity and consistency of items' content among Tehranian children.

Comparing the results with reported coefficients for American children (Ulrich, 2000), we found that the internal consistency reliability of the two subtests and GMQ was higher (.85, .88 and .91 respectively) among American. One of the factors affecting reliability is the group variability. Considering the concept of reliability in the classical test theory which views reliability as the proportion of true score's variance to observed score's variance, the more heterogeneous the group the higher the reliability score would be, because of the broad range of scores (Kubiszyn and Borich, 1990). When the standard deviation of the subtests' score for American children were compared with those of the participants in the present study, it was found that, in most of the age groups, the standard deviation for the scores of the American children were higher (Ulrich, 2000). It would be attributed to cultural and physical education content differences of two societies.

The estimation of the test-retest or stability reliability showed that the ICC coefficients for the object control and locomotion subtests and also GMQ were .65, .85 and .81 respectively. With regard to the criteria of Fleiss (1981) the resulted coefficients show the acceptable stability of TGMD-2 scores over time. The results indicate

that the scores of object control subtest are more stable than the scores of locomotion subtest. Although the comparison of the results to other studies is difficult due to the differences in the statistical analysis methods applied, Ulrich (2000) and Simons et al. (2007) also observed this difference among TD American children and mental retarded Flemish children, respectively (Simons et al., 2007; Ulrich, 2000). To justify this difference, we can refer to study of Lee et al. (2005). The researchers investigated the psychometric properties of TGMD-2 according to item response theory concepts. They found that the object control items were more difficult and more precise than those of locomotion, regardless of the group properties (Lee et al., 2005). The more precise the items on a test, it assesses a more extensive range of the ability it is assumed to assess, thus the test has a higher reliability (Baumgartner and Jackson, 1991). So, it can be said that the higher reliability of object control subtest relative to locomotion subtest is due to the more precise items.

The result showed that the high intra-rater reliability of TGMD-2. In all articles which were reviewed on TGMD reliability, the scoring or objectivity reliability of the test was reported as higher than .80. This is indicative of the clear and unambiguous administration and scoring of the test.

TGMD-2 validity: Although validity evidence for the TGMD-2 has been reported for typically developing children (Ulrich, 2000; Valentini, 2012; Van Waelvelde et al., 2004) and children with intellectual disabilities (Simons et al., 2007), research investigating the validity of TGMD-2 in different countries was insufficient. More importantly, the TGMD-2 had not been validated in Iran. Then we aimed to assess some aspect of the validity of the TGMD-2 for measuring fundamental movement skills in Tehranian children. The two-factor solution reported by Ulrich (2000) that consisted of 12 test items for representing locomotor skills and object control skills was tested with confirmatory factor analysis and the results compared to the those reported for American (Ulrich, 2000), Hong Kong Chinese (Valentini, 2012) and Brazilian (Van Waelvelde et al., 2004) children. The two-factor model of the TGMD-2 in this study did not result in exact fit as illustrated by significant chi-square value for the reason of sample size. The chi-square to degrees of freedom ratios, however, did not exceed 5 demonstrated that the model was acceptable. As displayed in Figure 1 each motor skill correlated satisfactory with the correspondent subtest and negative correlations were not observed between motor skills and subtests that supported the two factor model of TGMD-2 (Cronbach, 1989). In addition, the correlation between locomotion and object control was .91. This high correlation shows that these two factors both assess the gross motor skills. But, in some case where correlation coefficients show the

Table 5. Reported factor loading for two subtests of TGMD-2 in three validation studies.

Skill	Ulrich (2000) (N = 1,208)		Wong and Cheung (2010) (N = 626)		Valentini (2012) (N = 2,674)	
	Loc	OC	Loc	OC	Loc	OC
Run	.52		.64		.46	
Gallop	.66		.68		.71	
Hop	.70		.71		.66	
Leap	.49		.56		.53	
Jump	.59		.57		.53	
Slide	.69		.78		.55	
Strike		.75		.63		.69
Dribble		.61		.66		.56
Catch		.57		.64		.59
Kick		.65		.50		.75
Throw		.75		.66		.69
Roll		.67		.72		.45

relation between the subtests of a battery, it is expected that the coefficients would be acceptably high; too high coefficient means all the subtests measure the same ability not different aspects of a construct (Ulrich, 2000). Thus, the observed correlation may be indicative of the fact that the two subtests are not independent and the two-factor structure of TGMD-2 does not fit to the data. But the two factor structure of TGMD-2 provided a sufficient fit to the data as presented by small RMSEA and SRMR, and high GFI, AGFI, NNFI and CFI, based on suggested values. Therefore the two-factor structure of the TGMD-2 seems to be tenable to Tehranian children. In this study the resulted loading of the factors and fit indices were similar to those reported for typically developing children from the United States (Ulrich, 2000), Hong Kong (Wong and Cheung, 2010) and Brazil (Valentini, 2012), as indicated in Tables 5 and 6. It was also observed relatively high correlation (.71) between the two latent constructs (locomotor and object control) for Hong Kong Chinese children (Wong, 2006). However, confirmatory factor analysis of one-factor solution with 12 items did not show an acceptable fit ($\chi^2/df = 7.86$) compared to two-factor solution ($\chi^2/df = 3.40$), and the two-factor model represented a significantly better fit than the one-factor. This further confirmed the two-factor structure of the TGMD-2.

By comparing reported χ^2/df in Table 6 and the value obtained in this study it could be concluded that the factorial structure of TGMD-2 provided better fit to data for Chinese population than American and Iranian population.

Table 6. Reported goodness of fit indexes of factorial structure of TGMD-2.

	Ulrich (2000)	Wong and Cheung (2010)	Valentini (2012)
χ^2/df	5.29	3.40	*
GFI	.96	.95	.98
AGFI	.95	*	.95
RMSEA	*	.06	.06
CFI	*	.97	.88

*Not reported.

The literature shows that as typically developing children grow older, their motor skill performance improves due to physical maturation and experience, and older children demonstrate higher mastery level of gross motor skill performance than younger children (13,14). So it could be expected that a valid motor development test should be capable to detect developmental changes in performance. The resulted correlations between the subtests total score and age were significant and ranged from .66 to .79, which are in the high to very high range, according to MacEachron's (1982) rule of thumb interpretations (Ulrich, 2000). The results of analysis of variance, also, provided further support for developmental nature of the test, with both gender's performance improving with age. Such results suggest that TGMD-2 has sufficient validity to assess age-related development of gross motor skills in Tehranian children. The observed correlations range indicates that correlation between age and locomotor subtest's scores was smaller than those for object control. Such a thing was observed among American (Ulrich, 2000) and Dutch children (Niemeijer, 2007). TGMD psychometric properties examining showed that the object control items have higher degrees of difficulty than the locomotor items (Cole et al., 1991; Lee et al., 2005). From a developmental view, one expects children of 3 to 10 years of age to perform better on locomotor skills than ball skills, as shown in Table 4. So the better performance of children in locomotor subtest is indicative of the fact that their scores are close to maximum. Thus, due to the ceiling effect, locomotor score progresses with age, especially in higher age groups, more slowly than object control score. It shows that age differentiation capability of object control items is better than locomotor items.

Our results indicated sex related differences in object control subtest, with boys outperforming girls. These results were in line with previous studies conducted in other countries (Okely and Booth, 2004; Spessato et al., 2012; Wong, 2006), and in special population such as children with autism (Berkeley et al., 2001), mental retardation (Simons et al., 2007) and visual impairment

(Houwen et al., 2007). According to the motor development literature the proficiency of gross motor skills differ between males and females because of biological differences and sociocultural factors. Then, based on the results, TGMD is a valid test for distinguishing between the two groups.

Conclusion

Based on the findings it was concluded that TGMD-2 is a valid and reliable instrument for measuring gross motor development of Tehranian children. Although the fit indexes of two-factor structure were seen reasonable as a result of confirmatory factor analysis, the correlation between two factors, that is, locomotor and object control was high; this refutes the assumed underlying traits of the test. So it would be necessary to examine fit indexes of single-factor model and compare them to two-factor model's fit indexes. According to the Yun and Ulrich (2002), the validation process should not be limited to one approach, so it is also recommended to examine content and criterion validity in future studies (Yun and Ulrich, 2002). It should be noted that some object control items like striking are not widely used in Iranian traditional games and sports, this issues should be considered in future examining internal consistency and construct validity especially through exploratory factor analysis. In addition, because the sample was limited to Tehranian children caution is needed in generalization of the results to other Iranian population and in the interpretation of data. Therefore, a comprehensive validation study is suggested through selection of a representative sample that is stratified by age, gender and geographical region of Iran.

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

- Armitage P, Theodore C (1998). Encyclopedia of Biostatistics. John Wiley & Sons Ltd.
- Baumgartner TA, Jackson AS (1991). Measurement for evaluation in physical education and exercise science. 4th ed. Wm. C. Brown publishers.
- Berkeley SL, Zittel LL, Pitneey LV, Nichols SE (2001) Locomotor and object control skills of children diagnosed with autism. Adapted Physical Activity Q. 18(4):405-416.
- Burton AW, Miller DE (1998). Movement skill assessment. Human Kinetics.
- Burton AW, Rodgerson RW (2001). New perspectives on the assessment of movement and motor abilities. Adapted Physical Activity Q. 18:374-365.
- Chow SMK, Hsu Y, Henderson SE, Barnett AL, Lo SK (2006). The movement ABC: A cross-cultural comparison of preschool children from Hong Kong, Taiwan, and the USA. Adapted Physical Activity Q. 23:31-48.
- Cole E, Wood TM, Dunn JM (1991). Item response theory: A useful test theory for adapted physical education. Adapted Physical Activity Q. 8:317-332.
- Cools W, De Martelaer K, Samaey C, Andries C (2008). Movement skill assessment of typically developing preschool children: A review of seven movement skill assessment tools. J. Sports Sci. Med. 8:154-168
- Cronbach LJ (1989). Construct Validity after thirty years. In RL Linn (Ed.), Intelligence: Measurement, theory, and public policy. Bloomington, IL: University of Illinois pp.147-171
- Deflandre A, Lorant J, Gavarry O, Falgairette G (2001). Determinants of physical activity and physical and sports activities in French school children. Perceptual Motor Skills 92:399-411.
- Evaggelinou C, Tsigilis N, Papa A (2002). Construct validity of the test of gross motor development; A cross-validation approach. Adapted Physical Activity Q. 19:483-495.
- Fleiss JL (1981). Statistical methods for rates and proportions, 2nd Edition. New York: Wiley
- Gallahue DL, Ozmun JC (2006). Understanding motor development: Infant, children, adolescent, adult. 6th ed. McGraw-Hill International Edition.
- Goshi F, Demura S, Kasuga K, Sato S, Minami M (1999). Selection of effective tests of motor ability in preschool children based on pass or fail criteria: Examination of reliability, objectivity, and rate of passing. Perceptual Motor Skills 88:169-181.
- Hands BP (2002). How can we best measure fundamental movement skills? Health Sciences Conference Papers. http://researchonline.nd.edu.au/health_conference/5/
- Houwen S, Hartman E, Jonker L, Visscher C (2010). Reliability and validity of the TGMD-2 in primary-school-age children with visual impairments. Adapted Physical Activity Q. 27:143-159.
- Houwen S, Visscher C, Hartman E, Lemink KAPM (2007). Gross motor skills and sport participation of children with visual impairments. Res. Q. Exercise Sport. 78(1):16-23.
- Kubiszyn T, Borich G (1990). Educational testing and measurement. 3th ed. Harper Collins Publisher.
- Lee M, Zhu W, & Ulrich DA (2005). Many-Faceted Rasch calibration of TGMD-2. Internet Article.
- Linn LL (editor) (1988). Educational measurement. 3th ed. American Council on Education & National Council on Measurement in Education. Part I: Reliability.
- Merriman W, Barnet BE, Isenberg D (1995). A preliminary investigation of the relationship between language and gross motor skills in preschool children. Perceptual and Motor Skills 81:1211-1216.
- Morrow JR, Jackson Aw, Disch JG, Mood DP (2005). Measurement and evaluation in human performance. 3th ed. Human Kinetics.
- Niemeijer AS (2007). Neuromotor task training for children with developmental coordination disorder. Unpublished doctoral dissertation, University of Groningen, the Netherlands p.16.
- Okely AD, Booth ML (2004). Mastery of fundamental movement skills among children in New South Wales: Prevalence and sociodemographic distribution. J. Sci. Med. Sport. 7:358-372.
- Piek JP, Baynam GB, Barrett NC (2006). The relationship between fine and gross motor ability, self-perceptions and self-worth in children and adolescents. Hum. Movement Sci. 25:65-75.
- Reeves L, Broeder CE, Kennedy-Honeycutt L, East C (1999). Relationship of fitness and gross motor skills for five to six yr. old children. Perceptual Motor Skills 89:739-477.
- Saif AA (2004). Educational measurement, assessment and evaluation. 3th ed, Dowran publishing company (Persian publisher).
- Sarmad Z, Bazargan A, Hejazi E (2004). Research method in behavioral sciences. Aagah publishing company (Persian publisher).
- Simons J, Daly D, Theodorou F, Caron C, Simons J, Andoniadou E (2007). Validity and reliability of TGMD-2 in 7-10 yr. Flemish children with intellectual disability. Adapted Physical Activity Q. 25(1):71-82
- Skordilis EK, Douka A, Spartali I, Koutsouki, D. (2004). Depth perception of elementary school students with qualitatively evidenced locomotor impairments. Perceptual and Motor Skills 99:501-518.
- Spessato BC, Gabbard C, Valentini N, Rudisill M (2012). Gender

- differences in Brazilian children's fundamental movement skill performance. *Early Child Dev. Care* pp.1-8.
- Thomas JR, Nelson JK (1996). *Research methods in physical activity*. 3rd ed. Human Kinetics.
- Ulrich DA (2000). *Test of Gross Motor Development*. 2nd ed. Austin, TX: Pro-Ed.
- Valentini N (2012). Validity and Reliability of the TGMD-2 for Brazilian Children. *J. Motor Behav.* 44(4):275-280
- Van Waelvelde H, De Weerd W, De Cock P, Smits-Engelsman BCM. (2004). Aspects of the validity of movement assessment battery for children. *Hum. Movement Sci.* 23:49-60.
- Vira R, Ruadsepp L (2003). Psychological correlates of physical activity among seven through eight grades. *Hum. Movement Stud.* 44:501-517.
- Wong AKY (2006). *Construct Validity of the Test of Gross Motor Development-2*. Unpublished doctoral dissertation, Hong Kong Baptist University.
- Wong AKY, Cheung SY (2010). Confirmatory factor analysis of the Test of Gross Motor Development-2. *Meas. Phys. Educ. Exercise Sci.* 14:202-209.
- Yun J, Ulrich DA (2002). Estimating measurement validity: A tutorial. *Adapted Physical Activity Q.* 19:32-47.

Full Length Research Paper

Formative evaluation in Physical Education initial teacher training courses

Francisco Gonçalves

Travessa Comendador Seabra da Silva n.º 2263720-297 Oliveira de Azeméis, Portugal.

Received 7th September 2012; Accepted 21th November 2012; Published May 2014

The main objective of this study was to analyze the study plans of every institution offering teaching majors in Physical Education and Sports in Portugal, in order to determine the existence of formative evaluation issues in initial formation of future teachers. The sample was constituted by the curricula and study plans of every Physical Education and Sport teaching majors offered in every Portuguese university, as well as polytechnic Institute (graduate Schools); amounting to 29 higher education institutions. The returned data were 100%. Collected data were submitted for a descriptive statistical and a content analysis, using the NVivo software. Eleven (37.9%) of these institutions were higher schools for education, private colleges were also eleven and seven were (24.2%) state universities. Related with formative teaching approach, data suggest the existence of discrepancies between initial teacher training models for physical education nationwide, both in content and structure. There are, however, similar training mechanisms with defined standards, according to each Institution. There is also no evidence of consensus between the institutions, when it comes to the contents related with formative subjects. Students' success is proportional to the quality of the teaching and learning process.

Key words: Education, student, learning, evaluation, physical education, teaching.

INTRODUCTION

Current pedagogy, committed to the development of the student, bears in mind the construction of the man of tomorrow. The school takes part in the process of changing society, and as an institution it should open itself to the community (Ademoh, 2010). The school must revise its action process, organizing and planning tasks that place the student at their core (Hall and Smith, 2006). This paper will follow a few guidelines, namely the evaluation of students' learning, programs, methodologies

and strategies. It will also address the areas of teacher and institution evaluation. Evaluation has come to fill a progressively important role in all fields of human activity, and in areas so diversified as the educational. This investigation work will focus on the issue of evaluation on the educational sphere, in other words, the evaluation of learning and teaching.

The work of the teacher includes verifying and judging students' productivity, evaluating teaching results. Every

E-mail: franciscojmg@gmail.com. Tel: 00351 917 668 858.

Author(s) agree that this article remain permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

teacher should bear in mind that some students learn faster than others. Accordingly, each teacher should be able to recognize the differences among students, helping them overcome difficulties and move forward in their learning. Student productivity reflects the work developed along the process, being preponderant in school routine and assuming an educational character (Darling-Hammond, 2006).

While carrying out the whole evaluation process, teachers must bear in mind that, evaluation is collecting the necessary information for an improved performance (Arafat et al., 2010). It is an excellent regulator of the entire teaching and learning process. It is the conscience of the very own educational process (Fossey, 2007). The teacher should therefore have a basic knowledge concerning the regulation of the teaching and learning process through evaluation.

The purpose of evaluation is to collect, analyze and interpret the elements, assembled over time, relating to a teaching product or system, in order to answer the following question: to what extent are teaching goals achieved?

Evaluation should be perceived as a review process for previously defined goals. It derives from the very process of teaching and learning, working as a mechanism that confirms if the intended goals are effectively achieved (Wharton-Michael et al., 2006).

Evaluation allows an accurate and assertive identification of problems that, properly analyzed, can be solved bearing in mind student's success as the final product (Atilla, 2010). Problem analysis gives ways to a decision making focusing on pedagogical efficiency (Howell and Nolet, 2000).

Generally speaking, evaluation is connected with the collection of data, interpretation of such data and, ultimately, with consequent decision making (Jenkins and Curtin, 2006).

Below, some of the key factors that underlie the idea of general accountability will be summarized, when developing a learning improvement oriented evaluation (Gonçalves and Aranha, 2011):

1. Students should be actively involved in the evaluation process;
2. Feedback is fundamental and indispensable to the improvement of the process;
3. Evaluation should be used to regulate learning;
4. Students should develop self-assessment skills;
5. Evaluative information should be obtained by diversifying strategies, techniques and instruments;
6. Evaluation influences student motivation and self-esteem, affecting learning, which in turn influences evaluation (evaluation – learning – evaluation cycle).

For all this factors to come together, one needs to pay attention to some aspects: a progressively educational and constant evaluation should be developed; the quality

of absorbed information should prevail over quantity; evaluation should assume more relevance and classification should be relegated to a second plan; evaluation strategies, techniques and instruments should be diversified, maintaining a certain distance from emphatic test administration (Biswajeet and Saro, 2009).

Teachers should master formative evaluation, because this evaluation must accompany the entire teaching and learning process, identifying successful learning experiences and flawed ones, so that the latter can be overcome and students can achieve proficiency and success. Another vital aspect deals with the importance of feedback for student performance throughout formative evaluation, serving as a response to the performance data. This evaluation moment is the only way the teacher can orient the student, by interpreting his performance and letting him know what he is doing well and what he is doing wrong. This (formative) evaluation moment is the most accurate account of students' performance, and through it every step of the process is visualized and included in the classification of students' proficiency/performance (Smith, 2001). Obtaining a double feedback is therefore the purpose of formative evaluation. In the first place, it provides a feedback about the student, what stages he overcame and the difficulties he encountered. In the second place, it provides a feedback about the teacher, letting him know how program evolves and the obstacles it faces. By assessing how the teaching and learning process evolves, formative evaluation helps the teacher adjust his teaching tasks to student learning in the course of the classes (Weston, 2004).

Self-assessment is one of the pillars of formative evaluation, as it regulates the teaching and learning process. It enables the recognition of mistakes and helps finding alternative solutions.

To possess pedagogical knowledge teachers must detain a specific understanding of the subject-matters they are ascribed to (Stufflebeam and Wingate, 2005):

1. Subject-matter content knowledge;
2. Pedagogical content knowledge;
3. Curriculum content knowledge;
4. Knowledge of values and norms;
5. Procedural knowledge.

Different training models and programs have to follow certain epistemological assumptions and are subject to social, cultural and ideological constraints, distinctive in each professional occupation (Bozkurt, 2010). Different training programs and models assume different perspectives and follow different assumptions, creating different understandings of the teaching occupation and its role (Ann et al., 2006).

METHODS

The purpose of the present work is to analyze the study plans from

every Portuguese higher education institution with a teaching major in Physical Education and Sports, in order to examine the existence of evaluation related subject-matters. The returned data were 100%, from the entire population; and this assigns a high degree of reliability to the obtained results.

In the present study, statistical measures were done by *NVivo 9* (for a qualitative analysis), *Excel 10* (for a quantitative analysis), and *SPSS 20* (for a quantitative analysis). Computer based content analysis programs, like *NVivo* for content analyses, are an increasingly pressing requirement, resulting from the demand for a higher degree of coherence between the different stages of scientific work. In this sense, the credibility of disclosed results has to be more and more substantiated, so that new theories, based on the resulting conclusions, can be formulated.

A documental analysis was performed in order to withdraw conclusions based on the objectives proposed. This content analysis was performed on every curricular program and was based on the study plan of each institution.

The content analysis method allowed implementing a previously outlined instrumental plan. Its application was very effective in the sense that it allowed us to examine, without human interpretation or influence, how prospective teachers are trained. Therefore, unreliable surveys, questionnaires or interviews were completely excluded from this study. Through documental analysis, programs are facts, agreements and guidelines, nearly binding teaching guidelines for teachers and learning guidelines for students.

The typological of each course was analysed: degrees, namely in defined goals, curricular organization, pedagogical and scientific dimensions.

The following hypotheses were formulated:

H₁ – Study plans for teaching majors in Physical Education and Sports have subject-matters that comprise the teaching of evaluation.

H₂ – Different educational institutions have different curricular structures and study plans.

Portuguese higher education institutions reveal some asymmetries. It is important to note that: most of these institutions are located in the coastline and many of them are located in the North; there are 1763 undergraduate degrees: 647 in state universities, 58 in Catholic universities, 299 in private universities, 532 in public polytechnic institutes and 227 in private polytechnic institutes.

RESULTS

In this study, every Portuguese institution offering initial training in Physical Education teaching major was analyzed, amounting to twenty nine. There were 11 higher schools for education (37.9%), 11 private institutions and 7 state universities, corresponding to 24.2%. All subjects included in the initial training process were analyzed.

Private Institutions have the highest number of subjects, with an average of 44.5, followed by universities with 43.4. Higher schools for education mark the highest discrepancy, with an average of 40 subjects in their study plans. This evaluation field seems important to analyse because, teacher evaluation is a complex task. First and foremost, it requires a specific profile from the evaluator. In other works, not every teacher is capable of evaluating. The evaluator should be someone with specialized

knowledge, enormous sensibility, empathetic communication and analytical skills, teaching experience and a heightened sense of social responsibility (Stewart, 2007). He has to be an attentive professional, capable of listening, clarifying, encouraging and helping to find solutions, giving opinions and also negotiating, orienting, establishing criteria and assuming the risks attached to the consequences of his actions (Gonçalves and Aranha, 2011).

Higher schools for education have integrated teaching practices, that is, students are exposed to real teaching situations during their graduation and not only in the final internship stage of initial teacher training, usually observed in universities. The higher schools for education of Beja and Oporto possess the lowest number of evaluation related subjects (only one). On the other side of the scale, the higher schools for education of Coimbra and Leiria possess 8 and 5 subjects, respectively.

By analysing the institutions that offer Physical Education and Sports initial teacher training, it becomes clear that state universities largely surpass higher schools for education but they fall behind private institutions, which seem to give more attention to evaluation related subjects in their curricular plans. Increasing teaching quality implies raising the number of evaluation related subjects when training prospective teachers.

After a quantitative introduction, a qualitative approach was adopted, treating this work like a compilation, re-compilation, comparison and expression classification work, and favouring comprehension according to the expressive system in concern. The adoption of a systematic and theoretically oriented logic for the treatment of this information was preferred. The analysis comprised essentially three stages: categorization, categorization verification and the construction of type aggregations in each category.

Starting with the exclusively evaluation related subjects, several significant features were identified, namely some teaching methodology cases, type of contents taught and even the means to achieve established goals.

McNeill and Krajcik (2007) underline the lack of institutions training professional experts that are prepared for teaching evaluation.

Teacher's training influences teacher's performance substantially. Decision making should emanate from a sustained and solid theoretical basis – training – as opposed to an empirical process. Thus, teaching mistakes (including evaluation mistakes) should not take place exclusively in the work field. Some of them can be avoided with a simple reminder during graduation. It is therefore highly important that evaluation related training is perceived as a teaching priority, not only in theory but also in practice (Howell and Nolet, 2000).

Teachers need to be experts in some areas, highlighting the authors, planning, pedagogical intervention and evaluation. A good teacher's performance can only

be achieved if the teachers master the most basic aspects of their professional occupation (Paulsen, 2002). Teacher's training should be oriented in order to provide them suitable decision making tools concerning the need and the abilities of their students, especially when it comes to evaluation, considered by many to be the core of the entire educational process. Additionally, teachers should be able to pass to their students the elements that will help them strengthen, rectify and encourage their learning, improving efficiency and letting them assume an active part in the learning process (Sproule, 2002).

In teaching majors this should be a priority, since it is the most important element in the work of prospective teachers. For this reason, university needs to systematically review curricula, bearing in mind the latest and most significant advances in the investigational field, making sure that the new teachers are able to apply educational objectives (Mohr, 2004).

Pedagogical policies in higher education must include teacher's training, promoting their adjustment to social challenges and especially providing them problem solving skills, so they can gradually become real pedagogical mediators. For that to happen, universities should develop a closer relationship with employing entities, putting an end to a long-lasting inertia (Chall, 2002).

Conclusion

The teaching and learning process should underline the educational role of evaluation, the importance of student retention, the reinforcement of teacher and student's roles, and the articulation between the student's evaluation system and the evaluation of the teaching system.

Concerning the problem raised in this study, it can affirm that it is important to evaluate, because everyone has been or will one day be evaluated. Moreover, evaluation is too important to be neglected when training prospective teachers, especially considering its two purposes: regulation and classification.

When transmitting knowledge, theory always exceeds practice. Consequently, initial teaching training is clearly focused on the theoretical level, promoting knowledge but relegating action to a second plan. There should be a bigger concern with the correlation between knowledge and action, because *knowing*, *knowing how to do* and *doing* are closely interconnected.

Initial training can place the prospective teacher in a situation where he can anticipate his professional future, allowing him to develop better skills. It can also give him a sustainable and conscious understanding of the teaching and learning processes, as well as human development and motivations. Moreover, the performance of simulation trainings in probable professional situations enables the development of self-knowledge when faced with new realities.

Conflict of Interests

The author(s) have not declared any conflict of interests.

REFERENCES

- Ademoh N (2010). Evaluation of the tensile strength of foundry cores made with hybridized binder composed of Neem oil and Nigerian gum arabic. *Int. J. Phys. Sci.* 5 (5):557-563.
- Ann Van Eps M, Cooke M, Creedy D, Walker R (2006). Student Evaluations of a Year-Long Mentorship Program: A Quality Improvement Initiative. *Nurse Education Today* 26(6):519-524.
- Arafat O, Dimiyati K, Seeme F (2010). A comparative simulation of mobile WiMAX physical layer performance for Zero-Force (ZF) and Minimum Mean Square Error (MMSE) channel equalizers. *Int. J. Phys. Sci.* 5(16):2503-2515.
- Atilla P (2010). The effects of weighted rope training on muscle damage of basketball players. *Int. J. Phys. Sci.* 5(11):1787-1795.
- Biswajeet P, Saro L (2009). Landslide risk analysis using artificial neural network model focussing on different training sites. *Int. J. Phys. Sci.* 4(1):1-15.
- Bozkurt I (2010). Effects of exercises on bone mineral density of proximal femour region among athletes of different branches. *Int. J. Phys. Sci.* 5 (17):2705-2714.
- Chall J (2002). *The Academic Achievement Challenge: What Really Works in the Classroom?* New York: The Guilford Press.
- Darling-Hammond L (2006). Constructing 21st-century teacher education. *J. Teach. Educ.* 57 (3).
- Fossey R (2007). University Oversight of Professors' Teaching Activities: A Professor's Academic Freedom does not Mean Freedom from Institutional Regulation. *J. Pers. Eval. Educ.* 19:3-4.
- Gonçalves F, Aranha A (2011). Physical Education Evaluation Methods. *Int. J. Educ. Adm. Policy Stud.* 2(3).
- Hall T, Smith M (2006). Teacher Planning, Instruction and Reflection: What We Know About Teacher Cognitive Processes. *Quest* 58:424-442.
- Howell K, Nolet V (2000). *Curriculum-Based Evaluation: Teaching and Decision Making.* Belmont: CA. Wadsworth.
- Jenkins S, Curtin P (2006). Adapting job analysis methodology to improve evaluation practice. *Am. J. Eval.* 27(4):485.
- McNeill K, Krajcik J (2007). Scientific Explanations: Characterizing and Evaluating the Effects of Teachers' Instructional Practices on Student Learning. *J. Res. Sci. Teach.* – Wiley Periodicals, January, 1-26.
- Mohr D (2004). Curricular Issues in Physical Educations teacher Education: Which Foundational Courses do Physical Education Teacher Education Students Really need? Should More Time be Spent on Pedagogical Content Knowledge? Is the Curriculum Presented in the Best Order? *JOPERD – J. Physical Edu., Recreation & Dance* 10(1).
- Paulsen M (2002). Evaluating Teaching Performance: New Directions for Institutional Research. *Wiley Periodicals* 114:5-18.
- Smith R (2001). Formative Evaluation and the Scholarship of Teaching and Learning. *New Directions for Teaching and Learning* (88):51-62.
- Sproule R (2002). The Underdetermination of Instructor Performance by Data from the Student Evaluation of Teaching. *Economics of Education Review.* Pergamon – Elsevier 21:287-294.
- Stewart T (2007). Teachers and Leaders Evaluating Course Tasks Together. *English Language Teach. J.* 61(3):256-266.
- Stufflebeam D, Wingate L (2005). A self-assessment procedure for use in evaluation training. *Am. J. Eval.* 26(4):544-561.
- Weston T (2004). Formative Evaluation for Implementation: Evaluating Educational Technology Applications and Lessons. *American J. Evaluation: Elsevier – Science@Direct, Spring* 25(1):51-64.

Journal of Physical Education and Sport Management

Related Journals Published by Academic Journals

- Journal of Media and Communication Studies
- International Journal of English and Literature
- Philosophical Papers and Reviews
- Educational Research and Reviews
- Journal of African Studies and Development
- Philosophical Papers and Reviews

academicJournals