

## DETERMINING INTER-RATER RELIABILITY AND CONSTRUCT VALIDITY OF THE DANCE PERFORMANCE ASSESSMENT INSTRUMENT FOR ASSESSING CHILDREN IN TRADITIONAL DANCE

Eugenia Nikolaki<sup>1</sup>, Maria Koutsouba<sup>1</sup>, Fotini Venetsanou<sup>1</sup>,  
Giorgos Lykesas<sup>2</sup>, and Giorgos Fountzoulas<sup>1</sup>

<sup>1</sup>School of Physical Education and Sport Science, National and Kapodistrian University of Athens, Greece

<sup>2</sup>School of Physical Education and Sport Science, Aristotle University of Thessaloniki, Greece

Original scientific paper

### Abstract

The aim of the present study was to examine the reliability and validity of the Dance Performance Assessment Instrument (DPAI) for assessing dance performance of children in traditional dance, using the criteria of inter-rater reliability, potential ceiling and floor effects, and known groups' performance. For that purpose, 83 students, 11-12 years old ( $M=11.76+0.62$  years), were divided in three groups, according to their years of dance experience (novices, intermediates and experienced). Children were evaluated with the DPAI, after being video recorded during a traditional dance lesson, by two specially trained judges. To examine inter-rater reliability, the percentage of agreement and the Cohen's kappa coefficient were calculated on all the DPAI assessment criteria; whereas, the intraclass correlation coefficient (ICC) was computed in both assessment categories and the overall dance performance index. The validity was investigated using the Kruskal-Wallis test on the dance performance index of the three dance experience groups. According to the results, high levels of agreement between the two judges (80.7% - 100%) were obtained. Cohen's kappa coefficient ranged from 0.66 to 1 ( $p<.001$ ); while, ICC was 0.98 ( $p<.001$ ). Furthermore, statistically significant performance differences were revealed among the three groups ( $\chi^2 = 41.36-72.81$ ,  $p<.001$ ); whereas, ceiling effects were revealed only on two out of the thirty DPAI dance criteria. The current results provide support for the reliability and validity of the DPAI, indicating that it can serve as a suitable instrument for assessing dance performance of 11-12 years old children in traditional dance. However, further research, focusing on other validity and reliability criteria, is recommended, so that the technical adequacy of the DPAI as an assessment instrument in traditional dance is established.

**Key words:** evaluation, assessment, dance performance, technical adequacy.

### Introduction

Evaluation of learning and teaching is considered to be a process closely related to the quality of education (Bloom, Madaus, & Hastings, 1981). Even though there are many definitions of educational evaluation, most of them have a common thread. Generally, according to the literature (Kellaghan, Stufflebeam, & Wingate, 2003) educational evaluation is an ongoing systematic process of collecting, analyzing and interpreting various sources of information in order to determine the needs of students, the knowledge and skills gained by them, the fulfillment of predetermined educational goals, and the upgrading of the instructional practices. The pedagogical function of evaluation is supported by an extensive methodological research literature, indicating the positive effects of its application both on students' performance and motivation as well as on teachers' professional development (Barton, 2002; Bloom et al., 1981; Creemers & Kyriakides, 2008; Stiggins, 2002).

In the above educational context, evaluation in Physical Education (PE), as a school subject, should be a key component of its lesson planning. Providing appropriate diagnosis, intelligent guidance and feedback evaluation in PE serves as

an effective and robust pedagogical decision-making tool (Casbon & Spackman, 2005) for student learning and instructional process (Worthen & Sanders, 1973). In order to achieve this, evaluation made on the basis of assessment measurements, should be as objective as possible, while subjective evaluation carried out by teachers, primarily based on student's behavior, should be restricted (Matanin & Tannehill, 1994; Siedentop, Doutis, Tsangaridou, Ward & Rauschenbach, 1994). Therefore, evaluation must be conducted with effective techniques, using reliable and valid assessment instruments so that its results are credible.

Focusing particularly on traditional dance, as a part of the PE curriculum in several countries (Mattsson & Lundvall, 2015), textbooks and teaching materials provided for its instruction do not contain dance assessment instruments and guidelines for reliable dance performance assessment (Dania, 2009; Rhone, 2017; Slettum, Fox, Looney, & Jay, 2001). This lack of assessment instruments could be problematic as far as the provision of specific and on-going feedback of the subject's learning and teaching processes is concerned (Stiggins, 2002).

On the other hand, educational areas that receive little or no assessment data are considered less useful or necessary than those which receive frequent assessment (Miller, Linn, & Gronlund, 2013). In view of the aforementioned, it is self-evident that the importance of the subject of dance may be negatively affected by the absence of meaningful assessment, despite its positive effects in the holistic development of students supported by several research findings (Biber, 2016; Lykesas & Zachopoulou, 2006; Lykesas, Dania, Koutsouba, Nikolaki, & Tyrovola, 2017; Swindlehurst & Chapman, 2008; Venetsanou & Kambas, 2004).

According to the relevant literature, there is a limited number of dance assessment instruments (Dania, 2009; Rhone, 2017; Slettum et al., 2001), applicable for folk dancing, in PE. Instruments falling under this category are: a) the Folk-Dance Performance Checklist for Children (Slettum et al., 2001), b) the Folk Dancing Proficiency Rubric (Rhone, 2017), c) the Rubric for the Evaluation of the Greek Traditional Dance Performance (Pitsi, 2016), and d) the Dance Performance Assessment Instrument (DPAI) (Dania, 2009). Although those tools constitute an important step towards dance performance assessment, some of them lack documentation or evidence supporting their technical adequacy. For example, the Folk-Dance Performance Checklist for Children (Slettum et al., 2001) is not based on a theoretical research framework, related to the dance movement analysis. Moreover, there is no published evidence supporting the validity and reliability of the Dancing Proficiency Rubric (Rhone, 2017), whereas, for the Rubric for the Evaluation of the Greek Traditional Dance Performance (Pitsi, 2016), the judge's selection and training could be addressed.

In the light of the above-mentioned background, the DPAI (Dania, 2009) may be a suitable tool for the assessment of folk dancing performance. The DPAI consists of thirty criteria (specific movement and dance protocol standards) that assess particular elements of the structure and quality of dance movement (Koutsoumba, 2005). In a broad sense, the DPAI assess various motor skills, responses to visual and aural cues, and contributes to awareness of others by producing a mix of qualitative and quantitative data (Dania, 2009). The DPAI is based on a solid theoretical background, that of Laban Movement Analysis (LMA), developed by Rudolf Laban (Laban, 1975, 1980). LMA is a theoretical and experiential system for the observation, description, notation and interpretation of human movement. In LMA all movements of the whole body or the body parts are observed as a pattern of change that occurs in terms of four categories, namely Body (bodily dimensions of movement), Effort (movement dynamics), Space (the kinesphere) and Shape (shape change and shape quality), collectively referred to as BESS, since there is no movement

that does not evolve in space and time, bringing the weight of the body to flow (Maletic, 1987). On this basis, LMA makes the analysis of the structure and performance of dance movement in relation to: a) the parts of the body that are used, as well as the interrelationships within the body and the others, b) the directions and paths of the movement, c) the changing forms that the body creates in space during movement, and d) the qualitative changes in the energy of movement with respect to inner intention (Koutsouba, 2005), possible.

LMA is one of the most widely used and sufficiently substantiated systems that determine human movement behavior (Santos, 2013). It has been acknowledged and applied in numerous fields and settings dealt especially with movement creativity and computation, such as dance, PE, physical and mental therapy (Bernardet et al., 2019). In PE, LMA has been used as the basis for curriculum content in games, gymnastics, and dance, providing paradigmatic instruction and assessment, in order to make learning relatable, understandable, interesting and challenging (Langton, 2007). As far as the dance subject is concerned, various researchers have applied Laban's theory to the compilation of lists of criteria for the observation, and assessment of dance performance (Dania, 2009; Pitsi, 2016; Rhone, 2017; Sossin, 1987).

As far as the psychometrics of the DPAI are concerned, its construct validity as well as inter- and intra-rater reliability in adult dancers are sufficiently supported (Dania, 2009; Dania, Hatziharistos, Koutsouba, & Tyrovola, 2014). However, traditional dance is also taught during elementary school physical education classes. In view of this, in order for the DPAI to be used as a dance performance assessment instrument in elementary school children, its reliability and validity are needed to be established in the particular context.

In terms of reliability, assessment instruments that measure an observed behavior or phenomenon and are not self-reported should be tested for reliability among independent judges (Cohen, Manion, & Morrison, 2007). Moreover, one of the most criteria frequently used to test the construct validity of an instrument is that of known groups that measures the instrument's ability to discriminate two different groups of individuals, known to differ in a particular trait or variable of interest (Suri & Verma, 2010).

Based on the above, the aim of the present study was to investigate the technical adequacy of the DPAI for assessing dance performance of 11-12 year-old children in traditional dance. For that purpose, the following criteria were examined: a) inter-rater reliability, b) potential ceiling and floor effects, and (c)

differentiation of known groups' performance. The years of their participation in extracurricular traditional dance classes were used for classifying children into dance experience groups (novice, intermediate, experienced), since, according to the sport literature, experience can affect and differentiate sporting performance (McPherson, 2008).

## Methods

### Participants

A total of 83 students of two elementary schools in Naoussa, Northern Greece, (41 boys and 42 girls), aged 11-12 years ( $M = 11.76 \pm 0.62$ ) were volunteered to participate in the present study, following parental consent form. The students completed a short questionnaire about their participation in Greek traditional dance classes outside of school. Based on their answers, they were classified into one of three groups [novice ( $n = 27$ ), intermediate ( $n = 28$ ), experienced ( $n = 28$ )], according to their dance experience (novice: 0-1.5 years, intermediate: 1.5-3 years, experienced: >3 years). The study was approved by the Bioethics Committee of the School of Physical Education and Sport Science of the National and Kapodistrian University of Athens (1100/13-02-2019).

### Measurements

The DPAI is an observational assessment instrument based on the systematic monitoring of the dancers. Specifically, dancers are divided into pairs and are video-recorded with a digital camera during the performance of a dance. Two trained judges specialized in Labanotation system (a notation system for recording and analyzing human movement through the use of abstract symbols) (Dania, 2009; Hutchinson-Guest, 2005) are provided with the above videos and a video showing the assessment dance performed by a

model that is used as a reference point. The judges assess the dance performance of each dancer separately according to 30 criteria that regard specific elements of the structure and style of dance movement, grouped in six categories: a) Body, b) Time, c) Space, d) Weight, e) Shape and f) Flow. Indicative criteria for each assessment category of the DPAI are described in Table 1. It should be noted, however, that the DPAI is a flexible assessment instrument. Based on the basic structural and morphological analysis of each dance, some of the criteria included may be omitted, as unresponsive to its form (Dania, 2009).

The total duration of each video recording is 120 sec. For the assessment of dance performance, the interval recording method, 10 sec observation/10 sec recording, is used. Each category is observed and recorded separately, in the following order: Body Time, Space, Weight, Shape and Flow. Numerical values (0, 1) are assigned for each criterion (1=meets the performance standard, 0=does not meet the performance standard).

The maximum score for each category is 5, while the maximum total sum of all the categories, which represents the Total Index (TI) of dance performance, is 30. The scores of the two judges are summed per category and the TI. Agreeing with Dania (2009), in the present study, the "Karsilamas Aisse" dance was selected as the assessment dance, since it is characterized by isorhythmy (absolute repetition of the same rhythmic form during the process of melody), isometry (maintenance of the metric union during the entire music piece), symmetric sequence of movements in different directions in space, free use of space, pauses, turns, and constraint improvisation form structured according to the main dance phrase (Dania, 2009; IFMC Study Group for Folk Dance Terminology, 1974).

Table 1. Indicative criteria for each DPAI assessment category.

CATEGORIES	CRITERIA
BODY	Placement of the body's center of gravity in regard with its stance
TIME	Performance of movement in relation to the music tempo
SPACE	Spatial focus or attention during movement
WEIGHT	Shift of weight (ie., when changing place or support)
SHAPE	Directional and like movement
FLOW	Performance of pauses

### Procedure

The research study was conducted at the participants' schools during PE lessons. All measurements were conducted, according to the validation process of the instrument, by two experienced judges with well-established scientific background in Labanotation system, whose inter- and intra-rater reliability is supported by previous studies in adult dancers (Dania, 2009, 2013).

### Statistical analyses

Data was analyzed using IBM SPSS Statistics for Windows, Version 24. Descriptive statistics were computed for the six categories of the DPAI and the TI. To examine inter-rater reliability, the percentage agreement and the Cohen's kappa ( $k$ ) coefficient were used for all the assessment criteria of the DPAI; whereas, the Intra Class Correlation coefficient (ICC) was calculated on the summative score of each assessment category.

The percentage agreement is a commonly used index to assess inter-rater reliability for observational measurements (Graham, Milanowski, & Miller, 2012). To calculate the percentage of agreement the following formula was adopted: Percentage agreement = Total of agreements / Total of (agreements + disagreements) x 100 (Cohen et al., 2007) corrected as to its writing (Araujo & Born, 1985): Percentage agreement = Total of agreements / Total of (agreements + disagreements) x 100%. Percentage agreement from 75% to 90% demonstrates an acceptable level of agreement (Graham et al., 2008; Stemler, 2004). In the present study, the acceptable level of agreement was defined at  $\geq 80\%$ . The percentage agreement statistics is the most straight-forward and directly interpretable approach to establishing reliability; however, it does not take into account the agreement expected by chance. Thus, it may overestimate the level of true agreement among raters (Cohen et al., 2007).

In order to obtain a more complete picture of inter-rater agreement the Cohen's kappa coefficient that takes into account this element of chance (Cohen, et al., 2007) was also used for all assessment criteria. Kappa can range from 0 to +1. Values <0 indicate no agreement, 0.01-0.20 slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial and 0.81 to 1 almost total agreement (Landis & Koch, 1977).

## Results

Means and standard deviations of the dance performance of all groups for each category and the TI are depicted in Table 2.

Table 2. Means and standard deviations of participants in the DPAI by dance group.

CATEGORY	GROUPS		
	NOVICE n=27	INTERMEDIATE n=28	EXPERIENCED n=28
BODY	4.48 +1.76	5.82+0.86	8.50+1.75
TIME	2.81+4.19	9.29+1.46	9.82+0.47
SPACE	2.44+2.92	7.68+2.19	9.11+1.13
WEIGHT	2.81 +1.14	4.89+ 0.99	6.96 + 1.13
SHAPE	2.07 +0.47	4.11 + 2.54	7.75 +0.64
FLOW	1.56+2.02	5.14+1.99	8.43+1.91
TI	16.18 + 7.68	36.98 + 3.45	50.57 + 3.62

The results regarding inter-rater reliability indicated that the two judges had >94% agreement with each other in all assessment criteria. Kappa values were statistically significant (Body= 0.66-1, Time= 0.96-1, Space= 0.80-1, Weight= 0.97-1, Shape= 0.94-0.97, Flow= 98.8-1,  $p < .001$ ) for all assessment criteria (Table 3). Moreover, ICC values revealed excellent agreement between the judges for all assessment categories and TI (Body= 0.98, Time= 0.99, Space= 0.99, Weight= 0.99, Shape=0.98 Flow: 0.99, TI= 0.99,  $p < .001$ ) (Table 4).

Moreover, 95% confidence intervals were calculated as they reflect the accuracy of the estimation by demonstrating the range in which the true value is likely to be contained (DeVon et al., 2007).

Regarding the ICC, in the present study the ICC (2.1) model (two-way random effects, absolute agreement, single measures) (Shrout & Fleiss, 1979) was selected as the appropriate form for the reliability analysis. Furthermore, a 95% confidence interval was also computed. Regarding the interpretation of ICC, values up to 0.40 are considered low, from 0.40 to 0.59 fair, from 0.60 to 0.74 good and from 0.75 to 1 excellent (Cicchetti, 1994).

To determine the construct validity of the DPAI, using the criterion of known-groups, the Kruskal-Wallis test (Kruskal & Wallis, 1952) was conducted, since the homogeneity of variance assumption (Levene's test) was violated (Meyers, Gamst & Guarin, 2013). Participants' dance experience was defined as independent variable, while their dance performance (summative score of each category of the DPAI and the TI) as dependent. The significance level of the test was set to  $\alpha = .05$ . In addition, ceiling and floor effects were examined in all the assessment criteria of the DPAI, calculating the proportion of subjects in each group who achieved a 1 or 0 (Rodrigues, et al., 2013).

The results of Kruskal Wallis showed that years of dance experience significantly affected performance on each DPAI category and the TI (Body:  $\chi^2 = 46.41$ ; Time:  $\chi^2 = 41.36$ ; Space:  $\chi^2 = 47.19$ ; Weight:  $\chi^2 = 58.60$ ; Shape:  $\chi^2 = 53.55$ ; Flow:  $\chi^2 = 56.13$ ; TI:  $\chi^2 = 72.54$ ,  $p < 0.001$  in all cases). Mean rank dance performance score of the experienced group in each DPAI category and TI was higher than that of the intermediate and novice group (Table 5).

Table 3. Percentage agreement and Cohen's Kappa values per assessment criteria.

CRITERIA	%	k (95% CI)(p<.001)	CRITERIA	%	k (95% CI) (p<.001)
BODY 1	98.8	0.66 (0.04-1.28)	WEIGHT 1	100	1
BODY 2	96.4	0.90 (0.79-1.01)	WEIGHT 2	100	1
BODY 3	100	0.77(0.62-0.93)	WEIGHT 3	97.6	0.95(0.88-1.02)
BODY 4	97.6	0.94(0.86-1.02)	WEIGHT 4	100	1
BODY 5	100	1	WEIGHT 5	100	0.93(0.85-1.01)
TIME 1	100	1	SHAPE 1	96.4	0.93(0.85-1.01)
TIME 2	100	1	SHAPE 2	96.4	0.93(0.85-1.01)
TIME 3	100	1	SHAPE 3	94	0.88(0.77-0.88)
TIME 4	100	1	SHAPE 4	96.4	0.92(0.84-1.01)
TIME 5	96.4	0.91(0.82-1.01)	SHAPE 5	97.6	0.95(0.88-1.02)
SPACE 1	97.6	0.94 (0.86-1.02)	FLOW 1	100	1
SPACE 2	95.2	0.93(0.85-1.01)	FLOW 2	100	1
SPACE 3	97.6	0.95(0.88-1.02)	FLOW 3	98.8	0.97(0.93-1.02)
SPACE 4	100	1	FLOW 4	98.8	0.97(0.93-1.02)
SPACE 5	95.2	0.89 (0.78-0.99)	FLOW 5	100	1

Table 4. ICC values per assessment categories and TI.

CATEGORY	ICC (p<.001)	(95% CI)
BODY	0.986	0.978 - 0.991
TIME	0.998	0.997 - 0.999
SPACE	0.991	0.985 - 0.994
WEIGHT	0.994	0.991 - 0.996
SHAPE	0.979	0.967 - 0.986
FLOW	0.996	0.994 - 0.997
TI	0.998	0.997 - 0.998

Table 5. Mean rank performance score in DPAI categories and TI per dance group.

CATEGORY	MEAN RANKS		
	NOVICE GROUP (n=27)	INTERMEDIATE GROUP (n=28)	EXPERIENCED GROUP (n=28)
BODY	22.76	38.14	64.41
TIME	20.63	50.09	54.52
SPACE	17.54	47.43	60.16
WEIGHT	17.74	41.61	65.79
SHAPE	21.67	37.86	65.75
FLOW	17.56	42.32	65.25
TI	14.11	41.46	69.43

Finally, regarding floor and ceiling effects, it was revealed that only in two criteria of the Body category, the performance standards were met by

a high percent of participants of all dance groups (Table 6) causing ceiling effects. The rest 28 criteria presented neither floor or ceiling effects.

Table 6. Criteria presenting ceiling effect.

CATEGORY: BODY	GROUPS		
	NOVICE: n=27 %	INTERMEDIATE: n=28 %	EXPERT: n=28 %
Placement of the body's center in regard with its stance	100	100	96.4
Body contact with other dancers	74.1	100	100

## Discussion

The aim of the present study was to investigate the reliability and construct validity of the DPAI for assessing 11-12 year-aged-children's dance performance in traditional dance. The reliability of the DPAI was determined using the reliability

between two independent judges, while the validity was assessed using the known groups criterion. Possible ceiling and floor effects were also estimated. For this purpose, three independent groups of children, known to have different levels of dance experience, were used. According to our results, the technical adequacy of the DPAI for

assessing children's dance performance is sufficiently supported. In particular, in terms of reliability, high levels of percentage agreement (94% - 100%) in all categories of the DPAI were observed. Similar levels were also observed in the study of Dania (2009), who developed and validated the DPAI in adult dancers. Taking into consideration that the predefined percentage agreement was 80% and that values from 75% to 90% illustrate an acceptable level of agreement (Graham et al., 2012), the above results indicate that both judges actually provided the same information.

As far as Kappa values are concerned, all the criteria of each assessment category achieved substantial or almost total agreement. The relative lower values of kappa for the criteria Body 1 (0.66) and Body 3 (0.77), despite the high agreement of these criteria (98.8% and 100% respectively), can be attributed to the imbalance in the table's marginal totals either vertically or horizontally, which is known as first Cohen's kappa paradox (Feinstein & Cicchetti, 1990). Generally, the values of Kappa were found to be higher as compared to those (0.43 to 0.87,  $p < 0.05$ ) previously obtained by Dania (2009). In addition, they are considered sufficient to draw reliable conclusions, as they range from 0.60 and above (Fleiss, 1981; Landis & Koch, 1977).

Apart from the above, ICC values were found to be excellent for all the categories of the DPAI and the TI, indicating that the two judges had a high degree of agreement. Taking into account the specialization of the judges in Labanotation system and the high level of agreement achieved by their responses, the application of the DPAI with only a single judge would be of interest, since this alternative may be more practicable and less time-consuming.

Regarding construct validity of the DPAI, the results of Kruskal-Wallis test revealed that the years of dance experience had a statistically significant effect on children's performance in all DPAI categories and the TI. Especially focusing on TI, a statistically significant difference among the three groups was observed, with experienced group presenting the highest scores, followed by the intermediate and novice group. Consequently, it can be argued that the years of dance experience increase dance performance.

As a result, the ability of the DPAI to classify each student's TI according to the years of his/her dance experience (novice, intermediate, expert) is supported. Similar results, regarding the construct validity of the DPAI based on the known-groups criterion, were observed in the study of Dania (2009) for the validation of the instrument ( $TI \times 2 = 67.591$ ,  $p < .05$ ). Furthermore, the above results are in line with the existing sport literature, according to which, domain specific training experience has been indicated as one of the most important factors of skill acquisition for expert

performance (Coughlan, Williams, McRobert, & Ford, 2014; Ford, Hodges, & Williams, 2013). For example, studies in traditional dance and ballet in adult dancers have shown that expert dancers are superior to novice in using different cognitive, perceptual, and motor skills (Issartel, Gueugnon, & Marin, 2017; Lin, Chen, Sua, Wub, & Lin, 2014; Munzert, Müller, Joch, & Reiser, 2019; Sofianidis, Hatzitaki, Grouios, Johannsen, & Wing, 2012). Also, the higher performance of expert athletes compared to that of novice ones has been reported in several sports, such as rhythmic gymnastics, in junior gymnasts aged 11-12 (Kioumourtoglou, Derri, Mertzanidou, & Tzetzis, 1997; Zisi, Giannitsopoulou, Vassiliadou, Pollatou, & Kioumourtoglou, 2009), as well as in adult athletes in volleyball (Araujo, Afonso, & Mesquita, 2011), basketball (Robins, Davids, Bartlett, & Wheat, 2008), water polo (Kioumourtoglou, Kourtessis, Michalopoulou, & Derri, 1997), football (Diaz del Campo, Villora, Lopez, & Mitchell, 2011), tennis (Villar, García-González, Iglesias, Moreno, & Cervelló 2007) and badminton (Blomqvist, Luhtanen, & Laakso, 2000).

Finally, participants' performance in the majority of the assessment criteria (28 out of 30) was free of floor or ceiling effects, with only two criteria of the body category (criteria 1 and 5) presenting ceiling effects. One could suggest that these criteria should be removed, since they do not provide any information about participants' dance performance. Nevertheless, it was decided to keep them in the tool, since their ceiling effects should not be seen as a limitation of the instrument. In order to explain the above result, consideration needs to be given to the structure and style of the dance "Karsilamas Aisse" that was used as the assessment dance. As mentioned above, this dance is characterized by free choreographic structure, free use of space, turns, pauses and close improvisation and is based on the development of two symmetrical contrasting motifs (right-left or front-back). Moreover, it is performed by couples, with the partners facing but not touching each other. As a consequence, there is no physical contact. Furthermore, the moderate and steady tempo of the dance helps dancers maintain their body balance (Dania, 2009). Therefore, the bunching of scores at the upper level in those specific criteria was expected; however, it was decided to keep the minimum in the Body category, since they could potentially yield different scores in another assessment dance.

This study has certain limitations that should be addressed. Firstly, the participants come from a specific region. Thus, the sample does not reflect the general population and the findings of the study cannot be generalized. In addition, the predefined participants' age limits the generalizability to other populations. Future research should, therefore, examine the ability of the DPAI to assess dance performance among children of different school ages. Secondly, the selected dance for the assessment of dance

performance ("Karsilamas Aisse") holds back the interpretation of the findings to other dances. Further exploration regarding the application of the DPAI in different dance categories, types, styles, or genres of dance would be worthwhile. Nevertheless, and despite its limitations, the current study presents an assessment instrument, that of DPAI, with a solid theoretical base, detailed and objective criteria for the assessment of dance performance and acceptable validity and reliability in the field of traditional dance. From an experimental point of view, the findings of the present study offer new, potentially useful information that can be served as a basis for the knowledge regarding the assessment of dance performance.

## Conclusion

In the present study, the reliability and validity of the DPAI for the assessment of 11-12-year-old children's dance performance in traditional dance were examined. The results provided sufficient evidence that DPAI has acceptable inter-rater reliability and construct validity, depicting its ability to discriminate the performance of children according to their dance experience. Yet, further research examining other validity and reliability criteria is needed in order for the technical adequacy of the use DPAI as an assessment instrument of children's dance performance in traditional dance to be established.

## References

- Araujo, R., Afonso, J., & Mesquita, I. (2011). Procedural knowledge, decision-making and game performance analysis in female volleyball's attack according to the player's experience and competitive success. *International Journal of Performance Analysis in Sport*, 11(1):1-13.
- Araujo, J., & Born D. G. (1985). Calculating percentage agreement correctly but writing its formula incorrectly. *The Behavior Analyst* 8, 207-208.
- Barton, P. E. (2002). Staying on course in education reform. Princeton, NJ: Statistics & Research Division, Policy Information Center, Educational Testing Service.
- Bernardet, U., Alaoui, S. F., Studd, K., Bradley, B., Pasquier, P., & Schiphorst, T. (2019). Assessing the reliability of the Laban Movement Analysis system, *PLoS-ONE*, 14(6).
- Biber, K. (2016). The effects of folk-dance training on 5-6 Years children's physical and social development. *Journal of Education and Training Studies* 4(11), 213-226.
- Blomqvist, M., Luhtanen, P., & Laakso, L. (2000). Expert-novice differences in game performance and game understanding of youth badminton players. *European Journal of Physical Education*, 25 208-219.
- Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). Evaluation to improve learning. New York: McGraw-Hill.
- Casbon, C., & Spackman, L. S. (2005). Assessment for learning in physical education. Leeds: Coachwise/ BAALPE.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education (6th ed.)*. London and New York, NY: Routledge Falmer.
- Coughlan, E., Williams, A.M., McRobert, & Ford. P. (2014). A novel test of deliberate practice theory: how experts learn. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 40, 449-458.
- Creemers, B.P.M. & Kyriakides, L. (2008). *The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools*. London: Routledge.
- Dania, A., Hatziharistos, D., Koutsouba, M. & Tyrovola, V. (2014). Validity and reliability of scores from a dance assessment instrument: Its application on Greek folk dance. In Overby, L. Y. & B. Lepczyk (Eds.), *Dance Current Selected Research, Volume 8* (pp 123-158). Brooklyn, NY: AMS Press Inc.
- Dania, A. (2009). Development of a dance performance assessment instrument. Its application on Greek traditional dance. Unpublished Master Thesis. Department of Physical Education and Sport Science, National and Kapodistrian University of Athens, Greece.
- Dania, A. (2013). From symbols to movement. The effects of the Laban notation method on Greek traditional dance learning. Unpublished Doctoral Dissertation. Department of Physical Education and Sport Science, National and Kapodistrian University of Athens, Greece.
- DeVon, H.A., Block, M. E., Moyle-Wright, P., Ernst, D.M., Hayden, S. J., Lazzara, D. J., Savoy, S. M., & Kostas-Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing Scholarship*, 39, 155-164.
- Diaz del Campo, G. D. González Villora, S., Lopez, L. M. G & Mitchell, S. (2011). Differences in decision-making development between expert and novice invasion game players. *Perceptual and Motor Skills*, 112(3), 871-888.
- Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of Clinical Epidemiology*, 43(6), 543-549.
- Fliss, J.L. (1981). *Statistical methods for rates and proportions (2nd ed.)*. New York: John Wiley.
- Ford P. R., Hodges N. J., Williams A. M (2013). Creating champions: The development of expertise in sports. In S. Kaufman (Ed.), *Beyond talent or practice: The complexity of greatness* (pp.391-413). NY: Oxford University Press.
- Graham, M., Milanowski, A., & Miller, J. (2012). Measuring and promoting inter-rater agreement of teacher and principal performance ratings. Center for Educator Compensation and Reform.
- Hutchinson Guest, A. (2005). *Labanotation. The system of analysing and recording movement (4th ed.)*. New York and London: Routledge.
- IFMC Study Group for Folk Dance Terminology (1974). *Foundations for the analysis of the structure and form of folk*

- dance: a syllabus. *Yearbook for Folk Music*, 6, 115-135.
- Issartel, J., Gueugnon, M., & Marin, L. (2017). Understanding the impact of expertise in joint and solo-improvisation. *Frontiers in Psychology*, 8, 1078.
- Kellaghan, T., Stufflebeam, D.L., & Wingate, L.A. (2003). *International handbook of educational evaluation*. Dordrecht: Kluwer
- Koutsouba, M. (2005). *Notation of dance movement. The passage from the pre-history to history of dance*. Athens, Greece: Propompos.
- Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks on one-criterion variance analysis. *Journal of the American Statistical Association*, 47, 583-621.
- Laban, R. (1975). *Laban's principles of dance and movement notation*. (2nd ed.) R. Lange. London: MacDonald and Evans.
- Laban, R. (1980). *The mastery of movement*. 4th edition revised and enlarged by LissaUllmann. London: MacDonald and Evans.
- Landis, J. R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. doi:10.2307/2529310
- Langton, T. W. (2007). Applying Laban's movement framework in elementary physical education. *Journal of Physical Education, Recreation & Dance*, 78(1), 17-53.
- Lin, C.-W., Chen, S.-J., Sua, F.-C., Wub, H.-W., & Lin, C.-F. (2014). Differences of ballet turns (pirouette) performance between experienced and novice ballet dancers. *Research Quarterly for Exercise and Sport*, 85, 330-340.
- Lykesas, G., & Zachopoulou, E. (2006). Music and movement education as a form of motivation in teaching Greek traditional dances. *Perceptual and Motor Skills*, 102, 552-662.
- Lykesas, G., Dania, A., Koutsouba, M., Nikolaki, E. & Tyrovolas, V. (2017). The effectiveness of a music and movement program for traditional dance teaching on primary school students' intrinsic motivation and self-reported patterns of lesson participation. *Mediterranean Journal of Social Sciences*, 8(1), 227-236.
- Maletic, V. (1987). *Body, space, expression: The development of Rudolf Laban's movement and dance concept*. Berlin. New York. Amsterdam: Mouton de Gruyter.
- Matanin, M., & Tannehill, D. (1994). Assessment and grading in physical education. *Journal of Teaching in Physical Education*, 13, 395-405.
- Mattsson, T., & Lundvall, S. (2015). The position of dance in physical education. *Sport, Education and Society* 20(7): 855-871
- McDermott, P. A. (1988). Agreement among diagnosticians or observers: Its importance and determination. *Professional School Psychology*, 3, 225-240.
- McPherson S. L. (2008). Tactics: using knowledge to enhance performance. In: Farrow D, Baker J, MacMahon C, (ed.) *Developing sport expertise: researchers and coaches put theory into practice*, pp. 155-167. London: Routledge.
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research: Design and interpretation* (2nd ed.). Thousand Oaks, CA: Sage.
- Miller, M. D., Linn, R. L., Gronlund, N. E. (2013). *Measurement and assessment in teaching* (11th ed.). New York, NY: Pearson Education.
- Munzert, J., Müller, J., Joch, M. & Reiser, M. (2019). Specificity of postural control: comparing expert and intermediate dancers. *Journal of Motor Behavior*, 51, 259-271.
- Pitsi, A. (2016). *Investigation of the effects of different teaching styles of Greek traditional dances on dance performance and motivation*. Unpublished Doctoral Dissertation. Departments of Physical Education and Sport Science. Democritus University of Thrace and University of Thessaly, Greece.
- Rhone, J. (2017). The challenge and benefit of evaluating folk dancing quality. *General Music Today* 31(2), 10-15.
- Robins, M., Davids, K., Bartlett, R., & Wheat, J.S. (2008). Changes in compensatory variability as a function of task expertise and distance during basketball shooting. In Y. Kwon, J. Shim, J.K. Shim & I. Shin (Eds.), 26 *International Symposium on Biomechanics in Sports* (pp. 473-476). Seoul.
- Rodrigues, S. L. L., Rodrigues, R. C. M., São-João, T. M., Pavan, R. B. B., Padilha, K. M., & Gallani, M. (2013). Impact of the disease: acceptability, ceiling and floor effects and reliability of an instrument on heart failure. *Revista da Escola de Enfermagem da U S P* 47Rev. esc.enferm. USP 47(5), 1090-1097.
- Santos, L. C. (2013). *Laban Movement Analysis: A Bayesian computational approach to hierarchical motion analysis and learning*. Unpublished Doctoral Dissertation. Faculty of Science and Technology. University of Coimbra, Portugal
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420-428.
- Siedentop, D, Douthett, P., Tsangaridou, N., Ward, P., & Rauschenbach, J. (1994). Don't sweat gym! An analysis of curriculum and instruction. *Journal of Teaching in Physical Education*, 13,(4), 375-394.
- Slettum, B. S., Fox, C., Looney, M. A., Jay, D. M. (2001). Validity and reliability of a folk-dance performance checklist for children. *Measurement in Physical Education and Exercise Science*, 5(1), 35-55.
- Sofianidis G., Hatzitaki V., Grouios G., Johannsen L., & Wing A. (2012). Somatosensory driven interpersonal synchrony during rhythmic sway. *Human Movement Science*, 31(3), 553-566.
- Sossin, M. (1987). Reliability of the kestenberg movement profile. *Movement Studies: Observer Agreement*, 2, 23-28.
- Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating inter-rater reliability. *Practical Assessment, Research, and Evaluation*, 9(4).
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Suri, S., & Verma, N. (2010). Questionnaire validation made easy. *European Journal of Scientific Research*, 46(2), 172-178.

- Swindlehurst, G. & Chapman, A. (2008). Teaching dance: A framework for creativity, In J. Lavin (Ed.), Creative Approaches to Physical Education: Helping children to achieve their true potential (pp. 29-54). Abingdon: Routledge.
- Worthen, B. R., & Sanders, J. R. (1973). Educational evaluation: Theory and practice. Ohio, OH: Jones.
- Zisi, V., Giannitsopoulou, E., Vassiliadou, O., Pollatou, E., & Kioumourtzoglou, E. (2009). Performance level, abilities and psychological characteristics in young junior rhythmic gymnasts: the role of sport experience. *International Quarterly of Sport Science*, 4, 1-13.
- 

*Received: 20 December, 2020*

*Accepted: 20 December, 2021*

*Corresponding author:*

*Eugenia Nikolaki, Ph.D. candidate*

*School of Physical Education and Sport Science,*

*National and Kapodistrian University of Athens, Greece*

*E-mail: evinikolaki@gmail.com*