European Psychomotricity Journal, 2015; 7; 1, 3-10 SciPsyMot-Hellas EPJ

ORIGINAL ARTICLE

Reliability of the National Program for Playground Safety (NPPS) Report Card

·Christoforidis, Ch., Iliadi, S., Nikolaidou, Th., Giannakidou, D.
School of Physical Education and Sport Sciences, Democritus University of Thrace, Greece

Abstract
Playground injury has been recognized as an important issue in children's everyday life. In order to analyze the factors that lead to an accident, researchers usually use questionnaires or report cards that have been used in other accident research areas. Reliable assessment tools have to be established in playground safety field. In the current study two researchers evaluated the reliability of the USA's National Program for Playground Safety (NPPS) report card in 138 Greek public playgrounds throughout Greece. Cronbach's alpha (*a*) was applied in order to estimate the internal consistency of the form and for the total score was found to be .461. The inter-rater and the test-retest reliability of the evaluation form were determined by calculating the Kendall's tau-b and the McNemar's test. Coefficients were found .947 (p<.001) and .83 (p<.001) respectively. Also, the standard error of the measurement (SEM) was .443 and .385, the mean coefficient of variation (% CV) fluctuated between 0-23.6 (mean 2.02) and 0-20.2 (mean 1.23) respectively. Bland-Altman plots demonstrated that the vast majority of the total scores were within two standard deviations for the two raters and the two measurements. NPPS appears to be a reliable report card for evaluating playground safety.

Keywords: *Reliability, NPPS Report Card, Playground Safety Assessment, Public Health*

Introduction

In the last two decades there is a worldwide tendency about playground safety issues, because a large number of children receive daily emergency department care for injuries sustained on playground equipment (Hart, 2002). In comparison to adults, children's quality of life assessment is a more recent area of research. Relevant commissions give priority to research related to accidents occurred outdoors because modern way of living makes playground development due to risk management a necessity (Purtscher & Mayr, 1998).

Playground behavior may be significantly risky. Among the important causes of school-related injuries are those associated with the use of playground equipment (Boyce et al, 1984). During 1978 data from the National Electronic Injury Surveillance System (NEISS), a statistically representative sampling of hospitals with emergency departments, projected that about 155.500 playground-related injuries-almost equally divided between home and public playgrounds were treated in emergency departments nationwide. Also, there had been 118.000 such injuries in 1974 (NEISS, 1979). Ward (1987) reported that each year, 65.220 injuries (40% of all playground injuries) occur on climbers, and 42.150 injuries (22% of all playground injuries) occur on slides. However, risk of playground injury is much more likely when children use the equipment incorrectly (e.g., going down a slide head first). Similarly, Mack, Hudson and Thompson (1997) reported 156.040 injuries occurred on equipment designed for public use, concluding a significant increase of about 6%.

In response to the problem the National Program for Playground Safety of the USA (NPPS, 1996) developed an action plan that focuses on four areas of playground injury prevention: *supervision, age-appropriateness of equipment, suitable fall surfaces and equipment maintenance* (Thompson & Hudson, 1996). Although there's a great research interest (Cavnar et al, 2004; Laforest et al, 2001; Mack, Hudson and Thompson, 1997), no data have established the reliability of the assessment tools (Saelens et al, 2006). While several measures have been developed to assess children's overall physical activity, there is a lack of reliable measures to assess the locations in which children play (Veitch et al, 2009).

* Correspondence: Christos Christoforidis, Ph.D., School of Physical Education and Sports Science, University Campus, 69100, Komotini, Greece, e-mail: cchristo@phyed.duth.gr

In Greece, there is a progress in recording playground incidents (Christoforidis & Kambas, 2007; Petridou et al., 2002). Nevertheless, the level of danger that children face when they use playgrounds in Greece has not been rated yet. In order to achieve that, a specialized form for the identification of hazards is needed but has not been established yet. The purpose of the current research was to evaluate the reliability of the USA's National Program for Playground Safety (NPPS) report card.

Methods

Participants

One hundred thirty eight (n=138) public playgrounds placed all over the mainland of Greece (Peloponnesus, Attica, Macedonia, and Thrace) were assessed from April to November of 2011. Playgrounds were chosen according to the criteria of being in use by children of preschool and first school age (1-8 years old) and had enough of the typical equipment (swings, climbers, etc).

Assessment tool

Playgrounds were assessed on the base of the U.S.A. National Program for Playground Safety (NPPS) report card, which literally translated into Greek. Back-translation method was applied in order to check the correctness of the translation and any corrections adopted in the final text. NPPS consists of four categories of questions: 1) *Supervision potentials of the playground for the parent or/and care-givers* (included four questions about the presence of a parent or care-giver at the area of the playground during the playing activity, the capability of visual contact between parents and children during free playing or using part of the equipment, as well as the existence of a discernible rules-of-behavior panel). 2) *Age appropriate design of the equipment* (included six questions about the presence of separate areas for ages 1-5 and 5-8 years old, the existence of age specific equipment rules, sufficiency of entrances and exits for the equipment, appropriateness of protective guardrails, *design* preventing from climbing outside the equipment and on the supporting). 3) *Surface of the playground* (included five questions about appropriateness of the surface of the ground and the six foot use zone, the depth of loose fill, the covering of concrete footings and foreign objects absence). 4) *Maintenance of the equipment* (included eight questions about sufficiency of the playground equipment. Particularly, the questions referred to any sort of malfunction that could appear on the equipment, such as broken or missing parts, protruding bolts and noticeable gaps that could cause an injury, potential accident traps, rust, splinters and cracks and/or holes).

For each of the questions mentioned above, a positive or negative answer was asked (yes/no). For every positive answer, the playground gained one point. After finishing evaluation, the researcher had to totalize the score of each category by adding the points (positive answers) of the questions. Then, by adding the scores of the categories, the total evaluation score for the playground was obtained. Finally, there was a classification of the playground in terms of appropriateness for play according to the following scale: Safe playground: 23-20 points, safe environment, work needed on "no" checked areas: 19-16 points, potentially hazardous playground: 15-12 points, dangerous playground, start improvements: 11-8 points and do not allow play: 7 points and below.

Procedure

Before the study, intra-rater reliability had been evaluated for 10 examiners fulfilled the selection procedure (one week training process with final examination). For the evaluation

of intra-rater reliability, videotapes with the own observation and assessment of two playgrounds for each examiner, had been recorded. One month later, each examiner watched the same videos and scored again. The 3.1 intra-class correlation coefficient was used for statistical analysis and found to be excellent (ranged from R=.88 to R=.91).

The group of 10 examiners was separated into pairs and each pair evaluated approximately 28 playgrounds. Each pair visited and evaluated the playgrounds throughout Greece using the NPPS form. Initially, the area of the playground, as well as its actual address, was recorded. For the examination of the inter-rater reliability of the NPPS report card the two raters assessed each playground at the same time in the first visit. In order to examine the hypothesis that any rater should fill in the playground evaluation form in the same way, the inter-rater reliability of the evaluation form was estimated by using the Kendall's tau-b coefficient and the McNemar's test. Frequency analysis was applied to illustrate the scores of the two raters in detail (Table 1).

Table 1. Frequency analysis for the total playground scores of the two raters

	1st rater		2nd rater	
	Frequency	%	Frequency	%
Do not allow play	8	5.8	9	6.5
Dangerous	37	26.8	41	29.7
Possible dangerous	57	41.3	55	39.9
Safe enough	36	26.1	33	23.9
Total	138	100	138	100

For the total score of the form, the within subject variation was reported as standard error of the measurement (SEM) and was expressed as a coefficient of variation (% CV). Bland-Altman plot was used to display visual representation of the errors against true values by plotting the difference between the two raters of the inter-rater reliability procedure on total score of the report card [rater 1 minus rater 2 (R1-R2)] against the mean of the scores of the two raters. Statistical significance was set at $p < 0.05$. Results showed a total score's Kendall's tau-b=.947 (p<.001). SEM was found .443 and %CV fluctuated between 0 and 23.6 (mean 2.02). The Bland-Altman plot demonstrated that the vast majority of the total scores were within two standard deviations for the two raters (Figure 1).

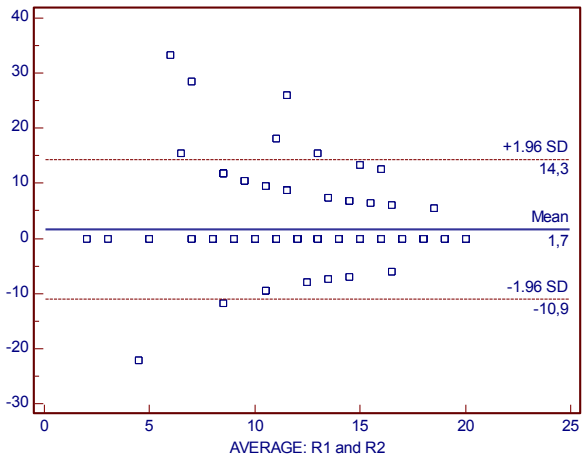

Figure 1. Bland-Altman plot for Rater 1 (R1) and Rater 2 (R2) assessments of the total score of the report card. Solid line represents mean difference between two raters. Dashed lines represent difference between two raters ± 1.96 S.D.

As far as concern the subcategories, *supervision's* Kendall's tau-b was equal to 1 (absolute agreement), *design's* Kendall's tau-b=.925 (p<.001), *surface's* Kendall's tau-b=.981 (p<.001) and *maintenance's*=.959 (p<.001). Frequency analysis was carried out to display the scores of the two raters. Results are shown in Table 2.

Table 2. Frequency analysis for the category scores of the two raters

	Supervision				Design				Surface				Maintenance			
	1st		2nd		1st		2nd		1st		2nd		1st		2nd	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
0	2	1.4	2	1.4	-	-	2	1.4	7	5.1	6	4.3	-	-	-	-
1	7	5.1	7	5.1	20	14.5	26	18.8	21	15.2	19	13.8	11	8	13	9.4
2	23	16.7	23	16.7	33	23.9	31	22.5	28	20.3	30	21.7	19	13.8	18	13.1
3	104	75.4	104	75.4	37	26.9	35	25.5	40	29	40	29	14	10.1	14	10.1
4	2	1.4	2	1.4	41	29.7	37	26.8	30	21.7	31	22.5	22	15.9	22	15.9
5	-	-	-	-	6	4.3	6	4.3	12	8.7	12	8.7	14	10.1	16	11.6
6	-	-	-	-	1	.7	1	.7	-	-	-	-	25	18.2	24	17.4
7	-	-	-	-	-	-	-	-	-	-	-	-	24	17.4	24	17.4
8	-	-	-	-	-	-	-	-	-	-	-	-	9	6.5	7	5.1
Total	138	100	138	100	138	100	138	100	138	100	138	100	138	100	138	100

To examine the test-retest reliability, each of the raters visited again, after three days, the randomly selected playgrounds (approximately 50% of the playgrounds that he/she was responsible for) and reassessed it by using the same NPPS report card.

Data analysis

For the evaluation of the internal consistency of the form, the Cronbach's alpha (*a*) for the total form score and for each category score were calculated. High values of Cronbach's *a* would indicate that some questions do not add reliability to the form and should be eliminated, whereas low Cronbach's *a* values could be an indicator that some questions do not actually match the total score of the category (Cronbach, 1951).

Finally, the consistency of the form, from one time to another, was assessed by estimating test-retest reliability of the report card. Particularly, to check the hypothesis that in any time the form should be filled in the same way, the test-retest reliability of the evaluation form was assessed by using the Kendall's tau-b coefficient and the McNemar's test. Frequency analysis was applied to illustrate the scores of the two assessments in detail. For the total score of the form, the within subject variation was reported as standard error of the measurement (SEM) and was expressed as a coefficient of variation (% CV). Bland-Altman plot was used to display visual representation of the errors against true values by plotting the difference between the first and the second assessment of the test-retest reliability procedure on total score of the report card [total score test minus total score retest (T-R)] against the mean of the first and the second assessment. Statistical significance was set at $p < 0.05$.

Results

Internal Consistency

For the examination of the internal consistency of the total form, the Cronbach's *a* for the total scores of the four subdivisions was found to be .461. The Cronbach's *a* if Item Deleted for *supervision* was .471, for *design* .453, for *surface* .238 and for *maintenance* .276.

The same procedure was also followed for each subdivision by estimating the Cronbach's *a* for the questions of each subdivision. For *Supervision* was .45, for *Design* .237, for *Surface* .485 and for *Maintenance* .711. Table 3 demonstrates the values of Cronbach's *a* of each subdivision, if Item Deleted.

Table 3. Cronbach's alpha if item deleted for Subdivisions

Supervision		**Surface**	
Question	Cronbach's Alpha if Item Deleted	Question	Cronbach's Alpha if Item Deleted
1	0.519	11	0.36
2	0.335	12	0.372
3	0.11	13	0.477
4	0.503	14	0.49
		15	0.428
Design		**Maintenance**	
Question	Cronbach's Alpha if Item Deleted	Question	Cronbach's Alpha if Item Deleted
5	0.425	16	0.612
6	0.237	17	0.6
7	0.43	18	0.705
8	0.059	19	0.645
9	0.023	20	0.73
10	0.113	21	0.675
		22	0.759
		23	0.687

Test-retest reliability

In order to quantify the test-retest reliability of the total form score the Kendall's tau-b was calculated. Results showed a total score .83 ($p<.001$). Frequency analysis was carried out in order to display the scores of the measurements. Results are shown in Table 4.

Table 4. Frequency analysis for the total playground scores of the two measurements

	Test		Retest	
	Frequency	%	Frequency	%
Do not allow play	8	5,8	8	5,8
Dangerous	37	26,8	38	27,5
Possible dangerous	57	41,3	57	41,3
Safe enough	36	26,1	35	25,4
Total	138	100,0	138	100,0

SEM was found .385 and %CV fluctuated between 0 and 20.2 (mean 1.23). The Bland-Altman plot demonstrated that the vast majority of the total scores were within two standard deviations for the two measurements (Figure 2).

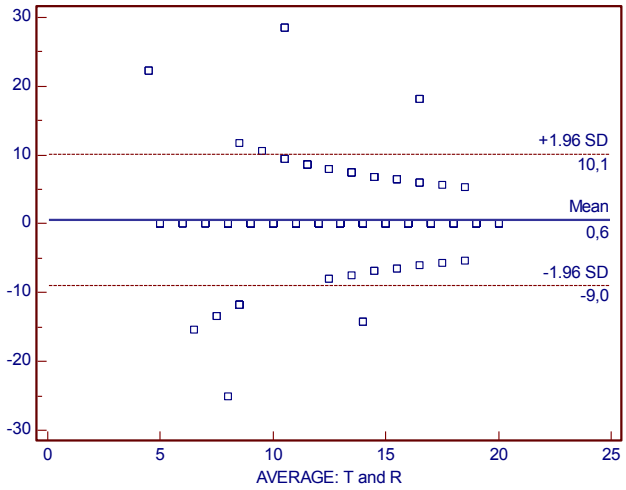

Figure 2. Bland-Altman plot for test (T) and retest (R) procedure for the total score of the report card. Solid line represents mean difference between the two measurements. Dashed lines represent difference between the two assessments ± 1.96 S.D.

As far as concern the subcategories, *supervision's* Kendall's tau-b was equal to 1 (absolute agreement), *design's* Kendall's tau-b=.935 (p<.001), *surface's* Kendall's tau-b=.933 (p<.001) and *maintenance's*=.935 (p<.001). Frequency analysis was carried out to display the scores of the two measurements. Results are shown in Table 5.

Table 5. Frequency analysis for the category scores of the two measurements

	Supervision				Design				Surface				Maintenance			
	Test		Retest		Test		Retest		Test		Retest		Test		Retest	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
0	2	1.4	2	1.4	-	-	-	-	7	5.1	7	5.1	-	-	-	-
1	7	5.1	7	5.1	20	14.5	19	13.8	21	15.2	21	15.2	11	8	11	8
2	23	16.7	23	16.7	33	23.9	33	23.9	28	20.3	28	20.3	19	13.8	19	13.8
3	104	75.4	104	75.4	37	26.8	38	27.5	40	29	40	29	14	10.1	14	10.1
4	2	1.4	2	1.4	41	29.8	41	29.8	30	21.7	30	21.7	22	15.9	22	15.9
5	-	-	-	-	6	4.3	6	4.3	12	8.7	12	8.7	14	10.1	14	10.1
6	-	-	-	-	1	.7	1	.7	-	-	-	-	25	18.2	25	18.1
7	-	-	-	-	-	-	-	-	-	-	-	-	24	17.4	24	17.4
8	-	-	-	-	-	-	-	-	-	-	-	-	9	6.5	9	6.5
Total	138	100	138	100	138	100	138	100	138	100	138	100	138	100	138	100

To check the test-retest reliability of each question of the form, McNemar's test was calculated. Results showed an absolute agreement for 21 of the 23 questions of the evaluation form, except two questions, the appropriateness of the surface of the ground (χ^2=64.2, p<.001) and presence of protruding bolts and head entrapments (χ^2=32.1, p<.001).

Discussion

The moderate Cronbach's alpha that came up from the results for the total form indicated a not very high consistency. This suggests that the four subdivisions were not examining the same risk factors, which agrees to the requirements of such an assessment form. On the other hand, if the Cronbach's alpha was even lower (<.2), it would be probably suggested that the form would not be able to assess the danger level of the playground in the

same way the subdivisions did (Cronbach, 1951). Furthermore, the Cronbach's alpha if item deleted revealed that all categories were important parts of the total form. Finally, comparing the Cronbach's alpha of each subdivision with the alphas if item deleted of each question, made obvious that every question contributes enough to the category total score so that it cannot be excluded.

Results indicated very high inter-rater reliability for the total form score, as well as for the subdivisions and for each question separately. These results reflect the similar way that the raters evaluated the playgrounds. This is very important for a form like that, because it can be used by different people who are going to assess playgrounds throughout a country. This will decrease significantly the time of conducting a full range study, since it is not necessary for a single researcher to make all assessments. The reliability is not in terms of 100% coincidence of the assessments by the different raters because some questions can probably be answered subjectively. This may suggest the use of special instruments to measure gaps or surfacing materials (CPSC, 2008) or more detailed directions and terminology (e.g. "what is missing part").

Results revealed a great reliability between the two different measures conducted by the same person within few days. This evident is not in agreement to the results of Veitch et al. (2008), which indicate a moderate reliability (ICC=.5-.8). This is possibly due to the methods followed by the two studies. In the earlier study, the parents were asked to record all the necessary information about outdoor play of children, and there was a consideration about the effectiveness and the accuracy of parent report as a subjective measure. For the purpose of the present study, though, an experienced researcher conducted both measurements. This probably ensured that the exact same conditions should result the exact same assessment of the risk factors, if the rater is previously informed and trained.

On the other hand, Veich et al. (2008) suggested that the conditions were not always the same just because all other factors were. Bad weather, for example, might keep children away from playgrounds, so as the participation in play was decreased. This indicates that even for less reliable questions of the assessment form of the present study, the problem is neither the form itself, nor the rater, but the different conditions from one measurement to another.

To provide a safer play environment, playgrounds must have adequate supervision, be maintained continually, and be equipped with age-appropriate equipment and resilient surfaces.

References

Boyce, W.T., Sobolewski, S., Sprunger, L.W., & Schaefer, C. (1984). Playground equipment injuries in a large urban school district. *American Journal of Public Health, (74)*, 984-986.

Bruininks, R., (1978). *The Bruininks-Oseretsky Test of Motor Proficiency: examiner's manual*. Circle Pines, MN: American Guidance Service.

Chalmers, D., Marshall, S., Langley, J., Evans, M., Brunton, C., Kelly, A.M., & Pickering, A. (1996). Height and surfacing as risk factors for injury in falls from playground equipment: a case-control study. *Injury Prevention*, *(2),* 98-104.

Christoforidis, C. & Kambas, A., 2007, Childhood injuries in Greek school environment. *International Journal of Injury Control and Safety Promotion*, *14*(4), 262-263.

Cronbach, L., 1951, Coefficient alpha and the internal structure of tests. *Psychometrica*, 16(3), 297-334.

CPSC, 2008, *Handbook for Public Playground Safety.* U.S. Consumer Product Safety Commission.

Friedman, L., 2005, *Measuring Reliability: the intraclass correlation coefficient*. Retrieved 1-15-2008 from http://www.na-mic.org/iki/images/2/2e/ICC_LectureFriedman2005.ppt.

Frost, J. & Sweeney, T., 1996, *Cause and prevention of playground injuries and litigation: Case studies*. Wheaton, MD: Association for Childhood Education International.

Hart, R., 2002, Containing children: some lessons on planning for play from New York City, *Environment & Urbanization*. 14(2), 135-148.

Helps, Y. & Pointer, S., 2006, *Child injury due to falls from playground equipment, Australia 2002-04*. NISU Briefing. Australian Institute of Health and Welfare.

Hopkins, WG. Measures of reliability in sports medicine and science. *Sports Med* 30:1-15, 2000.

Hudson, S., Mack, M. & Thompson, D., 2000, *How safe are America's playgrounds*? Cedar Falls, IA: National Program for Playground Safety.

Kraemer, H. C., 1982, *Kappa Coefficient*. In: Encyclopedia of Statistical Sciences, S. Kotz, and N. L. Johnson, eds. New York: John Wiley and Sons.

Laflamme, L. & Menckel, E., 1999, Injuries in Swedish during recesses: distribution and patterns. *Safety Science*, 33, 89-101.

Mack, R., Hudson, S. & Thompson, D., 1997, A descriptive analysis of children's playground injuries in the United States 1990-4. *Injury Prevention*, 3, 100-103.

Martin, J & Cooper, C., 2005, Playground safety in south western Sydney. *Journal of Paediatrics and Child Health*, 41, 587-591.

Mott, A., Evans, R., Rolfe, K., Potter, D., Kemp, K. & Sibert, J., 1994, Patterns of injuries to children on public playgrounds. *Archives of Disease in Childhood*. 71, 328-330.

Mowat, D., Wang, F., Pickett, W. & Brison, R., 1998, A case-control study of risk factors for playground injuries among children in Kingston and area. *Injury Prevention*. 4, 39-43.

National Electronic Injury Surveillance System (NEISS Descriptive Leaflet). 1979. Home Playground Equipment Injuries Treated in Hospital Emergency Rooms-1978. Tabulation of Data from NEISS, 1978, 1977, 1976, and 1975. CPSC, Washington, DC.

National Program for Playground Safety (1996). Safety Report Card, Retrieved from http://www.uni.edu/playground/research/report_card_form.pdf.

Norton, C., Nixon, J. & Sibert, J., 2004, Playground injuries to children. *Archives of Disease in Childhood*. 89, 103-108.

Petridou, E., Sibert, J., Dedoukou, X., Skalkidis, I., Trichopoulos, D. (2002). Injuries in public and private playgrounds: the relative contribution of structural, equipment and human factors. Acta Paediatr. 91, 691-697.

Saelens, B.E., Frank, L.D., Auffrey, C., Whitaker, R.C., Burdette, H.L., Colabianchi, N. (2006). Measuring Physical Environments of Parks and Playgrounds: EAPRS Instrument Development and Inter-Rater Reliability. *Journal of Physical Activity and Health*. 3 (1), 190-207.

Shrout, P. & Fleiss, J., 1979, Intraclass Correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(7), 420-428.

Thompson, D. & Hudson, S., 1996, *The National Action Plan for the Prevention of Playground Injuries*. Cedar Falls, IA: National Program for Playground Safety.

Tinsworth, D. & McDonald, J., 2001, *Special study: Injuries and deaths associated with children's playground equipment*. Washington, DC: Consumer Products Safety Commission.

Veitch, J., Salmon, J. & Ball, K., 2008, The validity and reliability of an instrument to assess children's outdoor play in various locations. *Journal of Science and Medicine in Sport (In Press,* doi:10.1016/j.jsams.2008.09.001).

Ward, A. (1987, April). Are playground injuries inevitable? *The Physician and Sports Medicine,* pp. 162–169.