Universal Journal of Psychology 3(2): 55-63, 2015
DOI: 10.13189/ujp.2015.030205 http://www.hrpub.org

Tool for Assessing Responsibility-based Education (TARE) 2.0: Instrument Revisions, Inter-rater Reliability, and Correlations between Observed Teaching Strategies and Student Behaviors

Amparo Escartí[1,*], Paul M. Wright[2], Carmina Pascual[3], Melchor Gutiérrez[4]

[1]Department of Social Psychology, University of Valencia, Spain
[2]Department of Kinesiology & Physical Education, Northern Illinois University, USA
[3]Department of Physical Education Pedagogy, University of Valencia, Spain
[4]Department of Educational & Developmental Psychology, University of Valencia, Spain

Abstract Although the original Tool for Assessing Responsibility-based Education (TARE) has proven useful in several studies, it has limitations. The three-fold purpose of this article is to present a revised version of the TARE including a new section to measure students´ behaviors, analyze the inter-rater reliability of the instrument, and assess the relationships between results of teacher and student observations. Data from 120 3-minute intervals of instructional time in physical education and general education lessons were analyzed. Intra-class Correlation Coefficient (ICC) in conjunction with Standard Error of Measurement (SEM) analyses was conducted to assess the inter-rater reliability of the teacher observation section and the student observation section of the TARE 2.0. Additionally, differential analysis, and Pearson correlation coefficients were carried out. Findings indicate the various categories in the teacher and student observation sections have a high degree of inter-rater reliability and that there are many significant positive correlations between the two.

Keywords Systematic Observation, Teaching Personal and Social Responsibility, Teacher Behavior, Youth Development, School-based Programs.

1. Introduction

Schools that provide a safe, positive learning environment and embrace a holistic view of child development can offer an excellent setting for initiatives intended to promote well-being among children [29]. Because children spend so many hours of their life in school, teachers and peers are powerful agents of socialization. Many schools try to accomplish this through a variety of programs [1]. In light of the potential of school based programs for prevention and well-being, some investigators express concern over the lack of evidence supporting these programs and the need to develop instruments to assess program quality [5].

There are numerous approaches to the assessment of school-based programs, such as questionnaires and interviews, but observation, in particular, is fundamental to classroom research. Observations are used to measure not that observations measure.

Observational methods are particularly useful because they measure, in real time, the behavior of the teacher, the students, and the interactions that occur in the classroom. Moreover, they have an advantage in that they can be used for professional development in school-based programs. For example, the feedback provided to teachers from observational methods can be a valuable element in their professional development [18]. At the same time, the systematic observation of teacher and student behaviors in the classroom can be used to monitor school-based intervention programs. For these reasons, in the past decade, scientists have placed renewed emphasis on developing standardized classroom observational measures with adequate reliability and validity [28].

The purpose of this article is to describe a revised version of the Tool for Assessing Responsibility-based Education (TARE) [33], which is an observational instrument to assess instruction aligned with the notion of teaching personal and social responsibility as reflected in Hellison´s [13] Teaching Personal and Social Responsibility model (TPSR).

TPSR model is a well-established instructional model that has been identified as an exemplary approach to promoting responsibility and self-efficacy in physical education and youth sport programs [22,27]. The TPSR model emphasizes the need to teach, through sports and physical activity, values and behaviors that can contribute to the positive

development of students. TPSR is generally described in terms of five responsibility levels or goals: (1) respect for the rights and feelings of others, (2) self-motivation, (3) self-direction, (4) caring, and (5) transfer/outside the gym. The first four levels can be enacted directly in a physical activity program, whereas the fifth level, transfer/outside the gym, relates to transferring the first four levels and associated behaviors to other settings, such as the classroom, playground, or home.

The postulates of the TPSR model align with the literature on Positive Youth Development and Resilience Theory, that consistently finds children, independent of their social status, who thrive tend to share a series of personal competencies like responsibility, self-control, autonomy, and social skills, as well as a secure source of support and caring from competent adults who support the learning of such competencies. Therefore, teachers are vital to the success of school-based youth development programs because they are facilitators in charge of organizing learning opportunities and guiding students through experiences that promote the programs' philosophy and values [20].

After decades of development and dissemination within the United States, TPSR has been adopted and implemented in many other countries, in Spain (e.g. ([8,25], New Zealand [12], Brazil [23], South Korea [16,24], and Canada [3].

The original instrument, TARE [33], consists of three main sections: 1) Observable Teaching Strategies; 2) Personal-Social Responsibility Themes; and 3) Student Responsibility. The teaching strategies assessed are: (1) Modeling Respect: Teacher models respectful communication. (2) Setting Expectations: Teacher explains or refers to explicit behavioral expectations during the program. (3) Opportunities for Success: Teacher structures lesson so that all students have the opportunity to successfully participate and be included regardless of individual differences. (4) Fostering Social Interaction: Teacher structures activities that foster positive social interaction. (5) Assigning Responsibility: Teacher assigns specific responsibilities that facilitate the organization of the program or a specific activity. (6) Leadership: Teacher allows students to lead or be in charge of a group. (7) Giving Choices and Voices: Teacher gives students a voice in the program. (8) Role in Assessment: Teacher allows students to have a formal role in evaluation. (9) Transfer: Teacher directly addresses the transfer of life skills or responsibilities from the lesson beyond the program.

After each 5-min interval, the appropriate codes on a scoring sheet are circled to indicate which strategies were observed during that interval. Results published by Wright and Craig [33] in the American school context as well as those published by Escartí, Gutiérrez, Pascual, and Wright [9] in the Spanish school context, indicate that both English and Spanish language versions of the teacher observation section of the TARE have satisfactory levels of reliability, based on inter-rater agreement, and validity, based on reviews by expert panels. The remaining two sections of the original TARE function more as holistic rubrics, allowing observers to provide overall ratings at the end of a class on various aspects of program implementation and student behavior. It should be noted that a self-assessment form for teachers, the TARE post-teaching reflection, has been published and already employed by several TPSR researchers and practitioners [14,9,15,34].

The original version of the TARE has proven useful in practice for purposes of program evaluation and teacher training. However, limitations exist regarding the tool's usefulness in the area of research. Regarding the time interval sampling section for observing teaching behaviors/strategies, the use of a binary coding system (observed vs. not observed) yields data that can only be analyzed using descriptive statistics such as frequencies and percentages. Data from the remaining two sections that focus on general themes and overall student responsibility have value for program evaluation and/or post-teaching reflection in that they document impressions and contextual information. Finally, the first and second authors of the current study noted in reviewing existing data that a 5-minute interval for time sampling appears to be more than is needed, unnecessarily restricting the number of data points for analysis. Therefore, the three-fold purpose of this article is to: a) present a revised version of the TARE, incorporating a new section to measure students´ behaviors in social settings; b) to analyze the inter-rater reliability of the new instrument; and c) to assess the relationships between results of teacher and student observations.

2. Materials and Methods

2.1. Participants and Setting

The current observational study was conducted at two public schools located in a working class town near the city of Valencia (Spain). One of the schools serves students in the elementary grades that come from a neighborhood of low socio-economic status. The other is a secondary school that serves students from a neighborhood of middle to low socio-economic status. The observed participants included two female teachers and their respective groups of students. Teacher A, in the elementary school, was a general education teacher. At the time of this study, she held a teaching degree in elementary education, and was a state employee with ten years of experience. Teacher B, in the secondary school, was a physical education teacher. At the time of this study, she held both a physical education and general education degree, and was a state employee with fifteen years of experience. Both teachers had participated in TPSR training and were supposed to be delivering that program. The students of Teacher A were 13 third-graders (4 girls and 9 boys). They were all eight or nine years old at the time and all were Spanish-born Caucasians. The students of Teacher B, were 22 seventh-graders (14 girls and 8 boys). They were all 12 or 13 years old at the time. The students in this group were also all Caucasians (15 Spanish-born and 7 immigrants born in South America).

2.2. Instrumentation

2.2.1. Creation of the TARE 2.0.

Several changes were made in developing the TARE 2.0. Firstly, a 5-point rating scale was introduced to replace the original binary scale used in the teacher observation section. The new scale allows observers to rate, within each interval, the degree of implementation regarding each strategy. On the revised scoring sheet the new Likert scale is: 0 (Absent), 1 (Weak), 2 (Moderate), 3 (Strong), and 4 (Very strong). Another change to this section included the reduction of the observation period from 5- to 3-minutes. Based on small scale pilot testing with existing video data, this shorter interval appeared sufficient to yield valid and reliable results. Finally, a major change to the original instrument involved the replacement of the two general rating sections with a new section for observing student interactions using the same time interval sampling methodology as the teacher observation section. This new section and its development are described more fully in the following paragraphs.

2.2.2. Development of the Student Observation Section

The process for developing the content for the student observation section began with a review of literature. The objective was to identify a number of discrete and observable student behaviors that aligned with salient outcomes and processes identified in the TPSR literature as well as the broader positive youth development. The intention was to develop a new section that would have a parallel structure and identical methodology with the revised teacher observation section. A decision was made to focus on behaviors that would occur in social settings and often involve group interactions rather than focusing on one individual at a time.

The first and second author went through several rounds of refinement and revision concerning the content of this section. This process was grounded in the extant literature but also in practice. For example, early versions of the new section were revised after using them to analyze segments of existing video data from the ongoing TPSR program. After several rounds of revision and the development of operational definitions informed by these practice sessions, a draft section was ready to field test. This field testing process involved the first and second author visiting a local elementary school and observing three different classes taught by three different teachers. The two observers independently rated student behaviors using 3-minute intervals, the operational definitions at that time, and the 5-point Likert rating scale used in the revised teacher observation section. The observed classes varied in subject matter (English Language, Physical Education, and Social Sciences). Class sizes ranged from 11-17 and all students were between the ages of eight and 10. After the observations were completed, reliability was assessed by calculating the percentage of inter-rater agreement. Although the level was acceptable, exceeding 80% overall, some final adaptations to the operational definitions were made based on the debriefing between the two authors.

The final version of the new section includes the following categories of student behavior: (1) Participation: Student is 'on task', i.e. following directions and participating in activities or tasks organized by the teacher. (2) Engagement: Student seems to have a high level of interest and motivation for the task or educational activity which could be evidenced in their level of active contribution. (3) Showing Respect: Student is actively showing respect to others, i.e. making eye contact, paying attention to others, or active listening. (4) Cooperation: Student demonstrates the social skills needed to work effectively with others in accomplishing a common task. (5) Encouraging Others: Student offers social support to others in proactive ways. (6) Helping Others: Student takes on helping roles. (7) Leading: Student takes on a leadership role with regard to an educational task. (8) Expressing Voice: Student makes suggestions, shares opinions, and/or reflects in ways that express their personality and individuality. (9) Asking for Help: Student seeks out assistance and asks for help from teacher or peers.

After completing a preliminary assessment of the new section for reliability, its validity was assessed. While the case for content validity rests in the direct alignment with the TPSR and youth development literature, a panel of experts was organized to evaluate the construct validity, i.e. the extent to which the content reflected the constructs of personal and social responsibility as framed by the TPSR model. All four authors were present when the expert panel was introduced to and asked to review the instrument. The panel consisted of seven individuals selected based on their experience related to the TPSR model, teaching children, and research methodology. The panel was comprised of two physical education teachers with an interest in TPSR who were enrolled in a master's degree program, one full-time general education teacher with basic knowledge of TPSR, two doctoral students with extensive knowledge of TPSR based in practical experience and academic study, one recent graduate from a doctoral program in counseling psychology who had done her doctoral research on TPSR, and a professor in sociology with substantial research experience related to TPSR.

The panel was provided with an introduction to the TARE including an explanation of changes being made to the original version. They were presented with the revised methodology and working definitions of both the teacher and student observation sections. Next, they were shown video segments as reference points to discuss the application of the operational definitions in practice. After these opportunities for open discussion and the chance to practice using the TARE 2.0 for video analysis, the members of the panel were asked to identify any categories, examples or content that appeared inconsistent with the constructs of personal and social responsibility as defined in the TPSR model. No inconsistencies were identified by any member of the panel. Finally, the panel was invited to give feedback on the revisions to the methodology and the structure of the new instrument. The majority of feedback was positive, however,

some small suggestions were offered that resulted in minor changes such as the names used for some codes in the rating scale and the inclusion of a comment section for contextual data to be re-introduced. All members were invited to provide additional follow up feedback via electronic mail within two weeks of the panel's original meeting; none did.

2.3. Procedure

After obtaining permission for the current study from the school district and consent from individual participants (active parental consent in the case of the students), video-recordings were taken of classes taught by Teacher A, the general education teacher, based in a classroom, and by Teacher B, the physical education teacher, based in her school´s gymnasium. Next, these video recorded lessons were analyzed independently by two observers, the first and second authors, who recorded their ratings on the data collection sheets. After establishing high levels of inter-rater reliability for both the teacher observation and student observation section of the TARE 2.0, data recorded by the first author were used for subsequent statistical analyses. In total, 120 distinct 3-minute intervals were analyzed and coded simultaneously using both the teacher and student observation sections. Sixty-one of these intervals were drawn from video-tape of Teacher A and her students engaged in four different lessons. The remaining 59 intervals were drawn from video-tape of Teacher B and her students distributed across five different lessons. Only complete intervals were analyzed, therefore, the overall data set represents 360 minutes of instructional time.

2.4. Data Analysis

The data were analyzed on three levels. Firstly, Intra-class Correlation Coefficient (ICC) in conjunction with Standard Error of Measurement (SEM) analyses was conducted to assess the inter-rater reliability of the teacher observation section and the student observation section of the TARE 2.0. As Weir [31] has stated, because of the relationship between the ICC and between-subjects variability, the heterogeneity of the subjects should be considered. A large ICC can mask poor trial-to-trial consistency when between-subjects variability is high. Conversely, a low ICC can be found even when trial-to-trial variability is low if the between-subjects variability is low. In this case, the homogeneity of the subjects means it will be difficult to differentiate between subjects even though the absolute measurement error is small. An examination of the SEM in conjunction with the ICC is therefore needed. Secondly, differential analysis between the observations of the two teachers and their pupils were carried out. Lastly, Pearson correlation coefficients were calculated to assess the relationships among data relating to teachers' strategies and their students' behavior. All data were analyzed using the SPSS software package.

3. Results

3.1. Inter-rater Reliability

As shown in Table 1, all the observed variables from the teacher observation and student observation sections of the TARE 2.0 have at least substantial correlation coefficients, with the exception of the variables Modeling Respect and Showing Respect. Landis and Koch [19] characterize values of reliability coefficients between 0.61 and 0.80 as "substantial" and those above 0.80 as "almost perfect". Nevertheless, considering ICC in conjunction with SEM, it can be concluded that the inter-rater reliability was high in all the variables, included Modeling Respect and Showing Respect, because of SEM value was in all cases very small [21,31]. Any comments and suggestions are welcomed so that we can constantly improve this template to satisfy all authors' research needs.

Table 1. Inter-rater reliability: Mean and SD for each observed variable, Intra-class Correlation Coefficients (ICC), p-values, Standard Error of Measurement (SEM), and p-values.

Variables	*M*	*SD*	*ICC*	*SEM*	*p*
Teacher Observation					
Modeling Respect	3.987	0.128	-0.023	0.019	.549
Setting Expectations	3.153	0.500	0.785	0.205	<.001
Opportunity for Success	2.756	0.978	0.821	0.375	<.001
Fostering Soc. Interaction	0.967	1.488	0.971	0.249	<.001
Assigning Manag. Tasks	0.384	0.909	0.908	0.262	<.001
Leadership	0.008	0.090	1.000	0.026	<.001
Giving Choices & Voices	0.967	1.434	0.946	0.324	<.001
Role in Assessment	0.207	0.815	0.994	0.062	<.001
Transfer	0.277	0.787	0.969	0.136	<.001
Student Observation					
Participation	3.112	0.571	0.771	0.240	<.001
Engagement	2.636	0.738	0.846	0.310	<.001
Showing Respect	3.884	0.340	0.433	0.168	.001
Cooperating with Peers	0.686	1.278	0.949	0.281	<.001
Encouraging Others	0.186	0.750	0.811	0.293	<.001
Helping Others	0.219	0.774	0.939	0.185	<.001
Leading	0.045	0.199	0.835	0.074	<.001
Expressing Voice	0.421	0.752	0.791	0.306	<.001
Asking for Help	0.174	0.394	0.794	0.159	<.001

3.2. Differences in Educational Strategies to Foster Responsibility Between General Teacher and PE Teacher

Table 2 shows the differences between the educational strategies to foster personal and social responsibility used by general education teacher and physical education teacher. Based on the data recorded with the teacher observation section of the TARE 2.0, there were statistically significant differences between the two teachers' ratings on Fostering Social Interaction, Assigning Tasks, and Giving Choices and Voices. In all these cases, it was the physical education teacher who used the various teaching strategies more often

with her students.

Table 2. Differences among educational strategies used by the General Education teacher and Physical Education teacher, registered by the teacher observation section of the TARE 2.0.

	Physic. Ed.	Gener. Ed.			
Variables	*M* (*SD*)	*M (SD)*	*t*	*df*	*p*
Modeling Respect	3.97 (0.18)	4.00 (0.01)	-1.41	119	.160
Setting Expectations	3.11 (0.32)	3.20 (0.63)	-0.94	119	.351
Opportun. for Success	2.90 (0.35)	2.65 (1.34)	1.42	119	.158
Fostering Soc. Inter.	0.13 (0.59)	1.88 (1.60)	-7.98	119	<.001
Assign. Manag. Tasks	0.25 (0.67)	0.58 (1.07)	-2.06	119	.041
Leadership	0.02 (0.13)	0.00 (0.03)	0.99	119	.323
Giv. Choices & Voices	0.26 (0.79)	1.77 (1.54)	-6.75	119	<.001
Role in Assessment	0.07 (0.51)	0.35 (1.02)	-1.94	119	.055
Transfer	0.31 (0.76)	0.25 (0.81)	0.42	119	.670

3.3. Differences between Students' Behavior in General Lessons and PE Lessons

Based on the data recorded with the student observation section of the TARE 2.0 (Table 3), students in the physical education classes received higher ratings than their counterparts in the general education classes on Participation, Engagement, Cooperating with Peers, Helping Others, Leading, Expressing Voice, and Asking for Help. On the contrary, there are significantly higher ratings among students in the general education classes on Showing Respect.

Table 3. Differences among pupils' behavior in General Education lessons and Physical Education lessons, registered through the student observation section of the TARE 2.0.

Variables	Physic. Ed.	Gener. Ed.			
Variables	*M (SD)*	*M (SD)*	*t*	*df*	*p*
Participation	2.85 (0.47)	3.32 (0.56)	-4.87	119	<.001
Engagement	2.33 (0.62)	2.98 (0.70)	-5.43	119	<.001
Showing Respect	3.97 (0.18)	3.77 (0.42)	3.38	119	.001
Cooperat. Peers	0.05 (0.28)	1.33 (1.54)	-6.73	119	<.001
Encourag. Others	0.18 (0.69)	0.28 (0.80)	-0.75	119	.452
Helping Others	0.10 (0.35)	0.38 (1.02)	-2.05	119	.043
Leading	0.01 (0.01)	0.08 (0.27)	-2.33	119	.021
Expressing Voice	0.23 (0.59)	0.68 (0.83)	-3.46	119	.001
Asking for Help	0.08 (0.27)	0.25 (0.47)	-2.38	119	.019

3.4. Correlations among the Strategies used by Teachers and their Pupils' Behavior

The results shown in Table 4 reflect the relationships between the strategies the teachers used to promote personal and social responsibility and the responsible behaviors their students demonstrated in general education and physical education classes separately. These results indicate there are fewer significant relationships between the educational strategies used by the general education teacher and the behavior of her students as compared to the number of significant relationships between the strategies used by the physical education teacher and her students' behaviors.

In the observations of the general education teacher and her students, there are significant positive correlations between the teaching strategy Setting Expectations and the student behavior Expressing Voice; between the teacher's delivery of Opportunity for Success and Engagement among her students; between Fostering Social Interaction as a teaching strategy used by the teacher and Cooperating with Peers, Helping Others and Asking for Help among her students; between the use of Assigning Tasks by the teacher and Cooperating with Peers, Helping Others and Expressing Voice among her students; between the teacher's use of Leadership as a strategy and Leading among her students; between Giving Choices and Voices by the teacher and Expressing Voice among her students; between the teacher's use of providing a Role in Assessment and Participation and Expressing Voice among her students; and finally, between the teachers use of the strategy promoting Transfer and Expressing Voice among her students.

With respect to the educational strategies used by the physical education teacher and the behavior of her students, there are positive and significant positive correlations between her use of Setting Expectations and her students' ratings for Participation and Cooperating with Peers; between her use of Opportunity for Success as a teaching strategy and Participation, Engagement, Cooperating with others and Helping Others among her students; between her ratings on Fostering Social Interaction and Participation, Engagement, Cooperating with Peers and Helping Others among her students; between her use of Assigning Tasks as a strategy and her students' ratings on Leading and Asking for Help; between her ratings on giving Leadership and ratings of Encouraging Others among her students; and between her use of Giving Choices and Voices as a strategy and Expressing Voice among her students. On the other hand, there are negative relationships between the physical education teacher's use of certain teaching strategies and some student behaviors. Her use of Modeling Respect as a teaching strategy was negatively correlated with Encouraging Others as a behavior among her students; her use of Giving Choices and Voices was negatively correlated with Participation and Cooperating with Others among her students; and finally, a negative correlation was also found between her use of Role in Assessment as a strategy and ratings of Participation among her students.

Table 4. Bivariate correlations among teachers' strategies to foster responsibility and their pupils' behavior.

		Students								
Teachers	Subject	Particip	Engagem	Showing Respect	Cooperat w. peers	Encourag others	Helping others	Leading	Express. Voice	Asking for Help
Modeling	Gen. Ed.	-.05	-.05	-.03	.03	.04	.05	.02	.07	.05
Respect	*P. Educ.*	*.07*	*-.01*	*-.07*	*.11*	*-.44***	*.05*	*.04*	*-.20*	*.07*
Setting	Gen. Ed.	-.21	-.02	.06	-.06	-.09	-.10	-.04	.56**	-.10
Expectations	*P. Educ.*	*.39***	*.23*	*.23*	*.42***	*.08*	*.22*	*-.10*	*-.16*	*.06*
Opportunity for	Gen. Ed.	.21	.37**	-.05	.05	.07	.08	.03	-.21	-.08
Success	*P. Educ.*	*.48***	*.54***	*-.11*	*.54***	*.14*	*.26**	*.08*	*,17*	*.19*
Fostering Social	Gen. Ed.	.07	-.12	.04	.55**	-.06	.66**	-.03	.20	.75**
Interact.	*P. Educ.*	*.34***	*.68***	*-.19*	*.62***	*.16*	*.32**	*.09*	*.06*	*.08*
Assigning Tasks	Gen. Ed.	.11	.08	.07	.63**	.05	.39**	-.05	.40**	.24
	P. Educ.	*.11*	*.04*	*-.03*	*-.14*	*-.17*	*-.09*	*.34***	*.06*	*.57***
Leadership	Gen. Ed.	.04	.14	.02	-.02	-.03	-.03	.99**	-.05	-.04
	P. Educ.	*-.07*	*.01*	*.07*	*-.11*	*.44***	*-.05*	*-.04*	*.21*	*-.07*
Giving Choices	Gen. Ed.	.10	.09	.06	-.06	-.08	-.09	-.04	.83**	-.10
and Voices	*P. Educ.*	*-.26**	*-.11*	*-.11*	*-.31**	*.17*	*-.06*	*.16*	*.55***	*.24*
Role in	Gen. Ed.	.31*	.14	.02	-.02	-.03	-.03	-.02	.39**	-.04
Assessment	*P. Educ.*	*-.31**	*-.16*	*-.04*	*-.17*	*.24*	*-.13*	*.01*	*.25*	*-.18*
Transfer	Gen. Ed.	-.23	-.11	.07	-.07	-.11	-.11	.11	.47**	-.12
	P. Educ.	*.05*	*-.14*	*.12*	*-.06*	*-.11*	*.05*	*-.09*	*-.15*	*-.16*

Note. * $p < .05$; ** $p < .01$; Gen. Ed. = General Education classes; *P. Educ.* = Physical Education classes.

4. Discussion and Conclusions

The purposes of this article were to present a revised version of the TARE including a new section to measure students´ behaviors, to analyze the inter-rater reliability of the new instrument, and to assess the relationships between results of teacher and student observations. This revised version of the TARE, named the TARE 2.0, is an observation instrument that, according to the results obtained in this study, measures, with a high degree of inter-rater reliability, teachers' and students' behaviors in the classroom to assess school based programs [5]. Moreover, the results indicate that there are numerous statistically significant relationships between teachers' use of responsibility-based teaching strategies and their students' enactment of responsible behaviors. As explained above, the original version of the TARE has proven useful in practice for teacher training and for purposes of program evaluation. The TARE 2.0 may prove more useful for researchers who aim to: a) evaluate the effects of teacher training on professional learning; b) evaluate the fidelity of the implementation of the TPSR model; c) evaluate the effects on student behavior; and d) asses the relationship between teacher training, implementation and student outcomes [15].

With regard to teacher training, it has been established that for programs to have the intended effects, they must be implemented effectively. Nevertheless, failure to implement effectively is a problem frequently documented in the literature [17]. Teachers are one of the key elements that influence implementation and therefore, their training is a point of growing interest in research on school based programs [7], including those based in TPSR [10]. The literature indicates the original TARE is a useful tool for promoting effective implementation and professional development when used in teacher training, both in the intensive training phase prior to implementation and the in-service training phase during implementation [6,10,15,32,34]. The use of the TARE in the intensive training phase can serve to introduce teachers to the core teaching strategies of the TPSR model. Training activities can allow teachers to identify, differentiate, and discuss the strategies. Moreover, they can practice these strategies through simulation techniques such as role-playing and receive individualized performance feedback. Throughout these activities, teachers have the opportunity to discuss different interpretations of the various strategies and their operational definitions with the trainer and their peers to solidify their understanding [15]. The TARE 2.0 can accomplish all of this in much the same way, but may be superior to the original instrument in that it yields more data points due to the shortened observation interval and provides more precise measurement due to the introduction of a five-point rating scale in place of the previous two-point scale.

In several published studies, the use of this self-assessment form in conjunction with feedback from direct TARE observation by a peer or trainer has proven effective in helping teachers to better understand areas of relative strength and weakness in their implementation and to continually improve their practice [6,10,15,32,34]. For example, Hemphill et al. [15] found this approach was useful in increasing teachers' understanding of the TPSR teaching strategies and the likelihood that they would apply them. This group noted that paired peer observations using this instrument increased the teachers' awareness of their

strengths and weaknesses regarding implementation and facilitated the teachers' learning, reflection, and discussions. In this way, the TARE framework provided a base for establishing a learning community [2].

Durlack and DuPre [7] argue that the evaluation of implementation is an absolute necessity in program evaluation as this is one of the principle reasons that implementation may fail. Therefore, new observational instruments are necessary to measure various aspects of implementation, such as dosage (how much of the original program has been delivered?) and quality (how well different program components have been conducted?). To strengthen implementation in TPSR programs, the TARE 2.0 can be used in initial training and then to monitor implementation as well as guide re-training as necessary. These functions can increase the fidelity of program implementation. For example, periodic checks with the TARE 2.0 allow for the identification of teachers who may be having problems with delivering certain aspects of the program. Moreover, the use of TARE 2.0 observations makes it possible to review compare and aggregate findings from multiple teachers to assess overall implementation to inform the improvement of ongoing group and individual training [30]. The original TARE was able to serve these same functions, but the TARE 2.0 has the advantage of yielding a greater quantity of more precise data to monitor and strengthen the teachers' delivery of the program as indicated by their used of the specific teaching strategies. Moreover, the TARE 2.0 takes a more rigorous look at student interactions which are crucial indicators that the program is being accepted and enacted by the students. The TARE 2.0 will be well-suited to make determinations regarding implementation fidelity (e.g., strong, moderate, or weak) because it yields a greater number of data points than the original instrument, includes rating scales with more gradations, and balances the focus on teacher and student behaviors.

Studies conducted on the use of the original TARE in teacher training and its impact [6,10,15,32,34], have helped to answer questions that are relevant for any intervention program e.g. what is effective professional development and how does it influence teachers' behaviors to support implementation? On this point, Garet, Porter, Desimone, Birman, and Suk Yoon [11] contend that more studies are required to determine the efficacy of different types of professional development activities. The results of studies like Hemphill et al. [15] that evaluate the use of the original TARE in teacher training help to establish links between this activity and its effects on teacher learning and practice. It also illuminates the extent to which the TARE framework contributes to the changes produced through program implementation. However, without more granular data related to students' behaviors, it is difficult to examine the relationships between teacher outcomes and student outcomes.

There are few studies on what teachers actually learn in their professional development and fewer still on what their students learn because of the changes that result in their teaching practice. Even though some studies in recent years have explored the relationships between the design and implementation of professional development and student learning outcomes [4,26], there is a lack of empirical evidence to explain how different forms of professional development may lead to different results in teacher behavior and student outcomes. Generally, the studies that have examined these relationships are based in self-report from the teachers as opposed to the direct observation of the professional development, the teachers' behavior, and the students' outcomes. By incorporating a new section to measure students' behaviors in social settings, the TARE 2.0 makes it possible to evaluate student outcomes in TPSR programs as well as program implementation. Therefore, this new instrument will allow researchers to examine the impact of professional development on teachers' behavior as it relates to student behavior.

Limitations in the current study include the relatively small number of participants overall and, in particular, the fact that only two teachers were involved. The small number of teachers makes difficult to generalize the results to teachers with different characteristics. Future studies should include a greater number of teachers as well as more variety in terms of gender, subject area taught, and grade level served. It should be noted that generalization is also limited by the fact that sampling was not random but based in convenience. However, despite these limitations, the current study and the instrument described make an important contribution to the literature. Based on the results of the current study, the TARE 2.0 appears to be an observational instrument that can be used to examine the effects of professional development as reflected in the behaviors of the teachers and their students with a high degree of inter-rater reliability. The TARE 2.0 also contributes to a gap in the literature related to evaluating the process of implementation as well as the effects of TPSR programs on students through observational methods to compliment the use of other methods such as questionnaires and interviews. While the development of the TARE instruments to date has occurred in the United States [33] and Spain [9], the TARE 2.0 and the findings presented here are relevant to a broader audience due to the international dissemination of the TPSR model [13]. In summary, the TARE 2.0 retains the value and functions of the original instrument, but is a key contribution in that it adds the new function of measuring student behavior and has been tailored to yield data of higher quality and quantity for rigorous research and evaluation studies related to the TPSR model or responsibility-based instruction in general.

Acknowledgements

This research project has been funded by the Spanish Ministry of Education (SEJ2007-67781) and the United

States Fulbright Commission.

REFERENCES

[1] American Academy of Pediatrics. Policy statement on school-based mental health services, Pediatrics, 113, 1839-1845, 2004.

[2] K.M. Armour & M. Yelling. Effective professional development for physical education teachers: The role of informal, collaborative learning, Journal of Teaching in Physical Education, 26, 177-200, 2007.

[3] S. Beaudoin, S. Using responsibility-based strategies to empower in-service physical education and health teachers to learn and implement TPSR, Agora for Physical Education and Sport, 14, 161-177, 2012.

[4] H. Borko. Professional development and teacher learning: Mapping the terrain, Educational Researcher, *33*, 8-15, 2004.

[5] R.F. Catalano, M.L. Berglund, J.A. Ryan, H.S. Lonczak, & J.D. Hawkins. Positive youth development in the United States: Research findings on evaluations of positive youth development programs, Prevention & Treatment, 5, 1-111, 2002.

[6] C. Coulson, C. Irwin, & P.M. Wright. Applying Hellison's Responsibility Model in a youth residential treatment facility: A practical inquiry project, Agora for Physical Education and Sport, 14, 38-54, 2012.

[7] J.A. Durlack & E.P. DuPre. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation, American Journal Community Psychology, 41, 327-350, 2008.

[8] A. Escartí, M. Gutiérrez, C. Pascual, & R. Llopis. Implementation of the Personal and Social Responsibility Model to improve self-efficacy during physical education classes for primary school children, International Journal of Psychology and Psychological Therapy, 10, 387-402, 2010.

[9] A. Escartí, M. Gutiérrez, C. Pascual, & P.M. Wright. Observation of the strategies that physical education teachers use to teach personal and social responsibility, Revista de Psicología del Deporte, 22, 159-166, 2013.

[10] A. Escartí, C. Pascual, M. Gutiérrez, D. Marín, M. Martínez, & S. Tarín. Applying the Teaching Personal and Social Responsibility Model (TPSR) in Spanish school context: Lesson learned, Agora for Physical Education and Sport, 14, 178-196, 2012.

[11] M.S. Garet, A.C. Porter, L. Desimone, B.F. Birman, & K. Suk Yoon. What makes professional development effective? Results from a national sample of teachers, American Educational Research Journal, 38, 915-945, 2001.

[12] B. Gordon. Teaching personal and social responsibility through secondary school physical education: The New Zealand experience, Agora for Physical Education and Sport, 14, 25-37, 2012.

[13] D. Hellison. Teaching responsibility through physical activity (3rd ed.). Champaign, IL: Human Kinetics, 2011.

[14] D. Hellison & P.M. Wright. Assessment and evaluation strategies. In, D. Hellison (Ed.), Teaching personal and social responsibility through physical activity, 3rd ed (pp. 161-182). Champaign, IL: Human Kinetics, 2011.

[15] M.A. Hemphill, T.J. Templin, & P.M. Wright. Implementation and outcomes of a responsibility-based continuing professional development protocol in physical education, Sport, Education, and Society, doi:10.1080/13573322.2012.761966, 2013.

[16] J. Jung & P.M. Wright. Application of Hellison's Responsibility Model in South Korea: A multiple case study of "at-risk" middle school students in physical education, Agora for Physical Education and Sport, 14, 140-160, 2012.

[17] K.A. Kealey, A.V. Peterson, M.A. Gaul, & K.T. Din. Teacher training as a behavior change process: Principles and results from a longitudinal study, Health Education & Behavior, 27, 64-81, 2000.

[18] J.M. Kuijpers, A.A. Houtveen, & T. Wubbels. An integrated professional development model for effective teaching, Teaching and Teacher Education, 26, 1687-1694, 2010.

[19] J.R. Landis G.G. & Koch. Measurement of observer agreement for categorical data, Biometrics, 33, 159-174, 1977.

[20] R.M. Lerner, E.M. Dowling, & P.M. Anderson. Positive youth development: Thriving as a basis of personhood and civil society, Applied Developmental Science, 7, 172-180, 2003.

[21] M.A. Looney. When is the intra-class correlation coefficient misleading? Measurement in Physical Education and Exercise Science, 4, 73-78, 2000.

[22] M. Metzler. *Instructional models for physical education*. Boston, MA: Allyn & Bacon, 2005.

[23] T.R. Montiero, R.K. Pick, & N.C. Valentini. Social and personal responsibility of child participants in an inclusive motor intervention program, Temas Sobre Desenvolvimento, 16, 202-214, 2008.

[24] D. Park, S. Nam, & C. Park. The experiences of the elementary school children of the TPSR Model applied in physical education lessons, Korean Journal of Sport Pedagogy, 14, 107-126, 2007.

[25] C. Pascual, A. Escartí, M. Gutiérrez, R. Llopis, D. Marín, & P.M. Wright. Exploring the implementation fidelity of a program designed to promote personal and social responsibility through physical education: A comparative case study, Research Quarterly for Exercise and Sport, 82, 499-511, 2011.

[26] W.R. Penuel, B.J. Fishman, R. Yamaguchy, & L.P. Gallagher. What makes professional effective? Strategies that foster curriculum implementation, American Educational Research Journal, 44, 921-958, 2007.

[27] A.J. Petitpas, A.E. Cornelius, J.L. Van Raalte, & T. Jones. A framework for planning youth sport programs that foster psychosocial development, The Sport Psychologist, 19, 63-80, 2005.

[28] R.C. Pianta, J. Belsky, R. Houts, F. Morrison, & the National Institute of Child Health and Human Development Early Child Care Research Network. Opportunities to learn in America's elementary classrooms, Science, 315, 1795-1796, 2007.

[29] E.P. Seligman, ER.M. Ernst, J. Gillham, K. Reivich, & M. Linkins. Positive education: Positive psychology and classroom interventions, Oxford Review of Education, 35, 293-311, 2009.

[30] J. Sheldon, A. Arbreton, & L. Hopkins. Investing in success: Key strategies for building quality in after-school programs, American Journal Community Psychology, 45, 394-404, 2010.

[31] J.P. Weir. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM, Journal of Strength and Conditioning Research, 19, 231-240, 2005.

[32] P.M. Wright. Offering a TPSR physical activity club to adolescent boys labeled "at risk" in partnership with a comprehensive community-based youth serving program, Agora for Physical Education and Sport, 14, 94-114, 2012.

[33] P.M. Wright & M.W. Craig. Tool for Assessing Responsibility-Based Education (TARE): Instrument development and reliability testing, Measurement in Physical Education and Exercise Science, 15, 1-16, 2011.

[34] P.M. Wright, B. Dyson, & T. Moten. Exploring the individualized experiences of participants in a responsibility-based youth development program, Agora for Physical Education and Sport, 14, 248-263, 2012.