



A Study of the Concurrent Validity between the Boxall Profile and the Strengths and Difficulties Questionnaire

Caroline Couture^{1a}, Paul Cooper^b and Egide Royer^c

^a *Université du Québec à Trois-Rivières, Quebec, Canada*

^b *University of Leicester, UK*

^c *Université Laval, Quebec, Canada*

The aim of the study is to establish the level of concurrent validity between the Boxall Profile, a diagnostic instrument used by teachers and teaching assistants in nurture groups, and the Strengths and Difficulties Questionnaire, a widely used screening instrument in the fields of education, mental health and social work. 202 children and adolescents attending nurture groups in England, aged 3-14 years, participated in the study. . These consisted of 142 boys and 60 girls and came from 25 schools in 8 LEAs. School staff completed the Boxall Profile and Strengths and Difficulties Questionnaire for all pupils. . The results show a high degree of concordance between the two instruments, with both measures appearing to identify similar behavioural characteristics in the same children. Scores in specific domains of the Boxall Profile are shown to predict performance on particular sub-scales of the Strengths and Difficulties Questionnaire. These preliminary findings support the validity claims of the Boxall Profile, indicating that it is a reliable tool for both diagnostic and research purposes.

Keywords: Boxall Profile, SDQ, validity, mental health, assessment

Introduction

Current concern with evidence-based practice in education often begs questions about what constitutes a valid and reliable assessment instrument. .Whilst it is not the contention of the current authors that positivistic, norm referenced assessment instruments provide the only basis for such evidence, it is clear that such instruments continue to play an important role in educational decision making. The current paper arises out of a study of the phenomenon of 'Nurture Groups' (NGs) which are currently enjoying something of a renaissance in the United Kingdom and are gaining popularity in other countries (e.g. Canada, New-Zealand, Malta). Nurture groups are special classes for children, usually located in mainstream primary schools, catering for between 10 and 12 students, and staffed by a teacher and a teaching assistant (TA). The

approach was developed by Marjorie Boxall, at the end of the 1960s, and is based on an understanding of early childhood development that draws on Bowlby's attachment theory for its conceptual coherence (Bennathan and Boxall 2000). A crucial feature of the NG approach is the *Boxall Profile* (BP), which is a detailed and rigorously trialed normative, diagnostic instrument (Bennathan and Boxall 1996; 1998) developed by Marjorie Boxall and her colleagues for use by teachers and TAs. Its purpose is to measure children's level of emotional and behavioural functioning, as well as to highlight specific targets for intervention within a child's individual functioning.

This instrument has been standardised for a population between 3 and 8 years (Bennathan and Boxall 1998) and has been in use in schools for more than twenty five years. The purpose of the present study is to establish the concurrent validity of the Boxall Profile with another questionnaire of a similar type. There is strong anecdotal evidence to suggest that the profile is highly regarded as a diagnostic and assessment instrument by a large number of teachers and educational psychologists who have used it (see Bennathan and Boxall 1998, 2000). This indicates a high level of context validity. The present study is an attempt to establish concurrent validity of the BP with a well validated screening questionnaire used in the field of behavioural difficulties, the teacher version of the *Strengths and Difficulties Questionnaire* (SDQ) (Goodman 1997; Goodman, Meltzer, and Bailey 1998; Goodman and Scott 1999). This will not only provide insight into the scientific credibility of the BP, but will also facilitate communication between practitioners and researchers who use these instruments.

Method

Sample description

As part of an ongoing study evaluating the effectiveness of nurture groups, information about 202 children and adolescents attending nurture groups in the United Kingdom has been collected (Cooper and Whitebread 2007). The children are aged between 3 years 11 months and 14 years 3 months ($X= 6.61$; $S.D. =1.90$), with 87.6% of the sample aged between 3 and 8 years, the range for which the BP has been standardised. 142 (70.3%) of the children are boys and 60 (29.7%) are girls.

The sample comes from 25 schools spread in 8 LEAs in the United Kingdom. The schools were selected on a convenient basis since all schools known as having a nurture group were invited to participate in the project. Once the school was enrolled, all children in the nurture group were included in our sample subject to consent from their teachers and parents. The majority (67.3%) of the children are at Stage 3 of the SEN Code of Practice., 28.1% at Stage 2, while a few (1 to 2%) at Stages 1,4, and 5. Only 2.5% of the children in the sample have a statement of special needs. English is not the native language of 15.7% of the children in the study and 8.8% in English according to their teachers.

¹ Corresponding author. Email address: Caroline.Couture1@UQTR.CA

Assessment tools

The teachers of the nurture groups were asked to complete the teacher version of the SDQ and the BP for each child in their classroom. Teachers had known children that they were asked to evaluate for a mean period of 2 terms prior to the evaluation. For a majority of the teachers, the BP is a tool that they regularly use for the assessment of progress of pupils in their nurture group. As a result, the teachers had sometimes already completed this questionnaire a few weeks before we asked them to do it again for the research project. To spare teachers the trouble of having to fill it a second time, we then accepted their filed results for the research purpose. However, all the SDQ were completed specifically for the research project since teachers do not usually use this questionnaire.

Boxall Profile: The profile is divided into two main sections. The first section, Developmental Strands, deals with developmental factors underpinning the individual's ability to engage effectively in the learning process. This section is divided into two clusters: Organisation of Experience and Internalisation of Controls. Each of these clusters comprises five sub-clusters that reflect the engagement of the child with the world and his or her level of personal development and awareness of others. The number of statements in each sub-cluster varies between 2 and 5, with a total of 34 statements in the section.

The second section, called Diagnostic Profile, deals with the child's behaviour characteristics that may inhibit or interfere with the child's social and academic performance. This section is divided into three clusters: 'self-limiting features' (2 sub-clusters), 'undeveloped behaviour' (3 sub-clusters) and 'unsupported development' (5 sub-clusters). These sub-clusters reflect 'lack of a normal thrust for growth', 'lack of inner resources to relate to others and engage at an age-appropriate level', and 'lack of early nurturing care' respectively. Like the first section, this section includes 34 statements split up into its 10 sub-clusters.

Preliminary internal reliability of the Boxall Profile has been investigated using data of the present study. Cronbach's alpha for each scale are shown in Table 1.

Strengths and Difficulties Questionnaire: The SDQ is a 25-item behaviour-screening questionnaire that measures 5 subscales, namely hyperactivity, conduct problems, emotional symptoms, peer problems and prosocial behaviour. It has been found to produce results consistent with more established behaviour rating scales, such as Achenbach's Child Behaviour Checklist, and Rutter's Child Behaviour Rating Scale (Goodman 1999). An extension to the questionnaire includes an impact supplement which seeks to reveal teacher, parent and child perceptions of the level of distress and social impairment associated with the symptoms revealed by the checklist (Goodman 1999). The psychometric properties of the SDQ can be found in Goodman and Scott (1999).

Table 1 Cronbach's alpha for the Boxall Profile scales

	<i>N</i> of items	α
Organisation of experience	5	.87
Internalisation of controls	5	.83
Self-limiting features	2	.24
Undeveloped behaviour	3	.51
Unsupported development	5	.83

Results

Discriminating Between Normal and Abnormal samples

After checking the normality of the distributions of the scores on each subscale of the SDQ and the BP, we performed independent samples T-test on the BP data to see if children falling in the normal and abnormal bandings of the SDQ were evaluated differently on the BP clusters. Since children in our sample receive particular provision in school and false positives were unlikely, we included the children falling in the borderline and abnormal ranges for this analysis. Results shown in Table 2 reveal that for four clusters out of five, normal and abnormal children, as assessed by the SDQ, obtained scores that are significantly different on the BP. Only the Undeveloped Behaviour cluster doesn't differentiate significantly between children who are in the normal range, following the SDQ rating, from those in the abnormal range.

The second analysis performed was the Pearson correlation matrix of the 6 scales of the SDQ with the five clusters of the BP. Table 3 presents this correlation matrix. We can see that each cluster of the BP is related in a significant manner to at least 3 scales of the SDQ and that all of them relate significantly to the total SDQ scale.

Developmental Strands: The cluster named "organisation of experience" comprises indicators that show the extent to which a child "is organised, attentive and interested", and the extent to which s/he is 'involved purposefully and constructively in events, people and ideas' (Bennathan and Boxall 1998, p.10). In addition to the total scale of the SDQ, this cluster is related significantly with the prosocial scale ($r=.389$, $p<.0005$), the peer problems scale ($r=-.297$, $p<.0005$), and the hyperactivity scale ($r=-.291$, $p<.0005$). These results suggest that children with a better organisation of experience also display better prosocial skills, less peer problems and less hyperactivity symptoms.

Table 2 T-test on the Boxall Profile between normal and abnormal SDQ scores

	Normal (N= 14)		Abnormal (N =170)		t (p)
	X	s.d.	X	s.d.	
Organisation of experience	47.57	14.06	38.61	13.06	2.45 (.015)
Internalisation of controls	46.07	11.16	33.82	10.35	4.23 (<.0005)
Self-limiting features	6.86	3.74	10.75	4.87	2.914 (.004)
Undeveloped behaviour	10.29	8.16	14.42	8.05	1.845 (.067)
Unsupported development	21.57	16.26	31.48	15.59	2.278 (.024)

BP – SDQ Correlations

Correlations between the five sub-clusters of the “organisation of experience” cluster and the five SDQ scales allow us to determine which sub-cluster has the strongest relationship with the SDQ. The strongest relationship in the sub-cluster “gives purposeful attention” is with the hyperactivity scale ($r=-.462$, $p<.0005$). “Participates constructively” ($r=.407$, $p<.0005$) and “connects up experiences” ($r=.269$, $p<.0005$) relate most strongly to the prosocial scale. Finally, it is with the Peer problem scale that “shows insightful involvement” ($r=-.382$, $p<.0005$) and “engages cognitively with peers” ($r=-.305$, $p<.0005$) are most strongly related. Within this cluster, the sub-cluster that relates most strongly with the SDQ Total scale is “gives purposeful attention” ($r=-.361$, $p<.0005$).

The cluster “internalisation of controls” comprises indicators that describe the extent to which a child “is emotionally secure, makes constructive, adaptive relationships, is able to co-operate with others, and has internalised the controls necessary for social functioning” (Bennathan and Boxall 1998, p.12). This cluster is significantly related to most of the SDQ subscales. The strongest relationship is with the prosocial scale ($r=.523$, $p<.0005$) and the weakest with the emotional symptoms scale ($r=.161$, $p=.029$). The data suggests that children who have a better internalisation of controls tend to have better prosocial skills, less conduct problems, less hyperactivity symptoms, less peer problems, but more emotional symptoms. This last relationship (even if it is a weak one) suggests that a better internalisation of controls could sometimes reflect too much internalisation, and the development of symptoms associated with this.

Table 3 Pearson Correlation Matrix between SDQ subscales and BP clusters

	SDQ 1	SDQ 2	SDQ 3	SDQ 4	SDQ 5	SDQ 6	BP 1	BP 2	BP 3	BP 4	BP 5
SDQ 1	1.000	-.166 .018	-.168 .017	.261 <.0005	.181 .010	.388 <.0005	-.016 .828	.161 .029	.274 <.0005	-.070 .348	-.055 .462
SDQ 2		1.000	.445 <.0005	.154 .028	-.526 <.0005	.639 <.0005	-.079 .283	-.464 <.0005	.084 .260	.304 <.0005	.577 <.0005
SDQ 3			1.000	.157 .025	-.358 <.0005	.653 <.0005	-.291 <.0005	-.449 <.0005	.105 .157	.404 <.0005	.200 .007
SDQ 4				1.000	-.256 <.0005	.63 <.0005	-.297 <.0005	-.208 .005	.192 .009	.168 .023	.029 .700
SDQ 5					1.000	-.418 <.0005	.389 <.0005	.523 <.0005	-.137 .065	-.090 .226	-.240 .001
SDQ 6						1.000	-.291 <.0005	-.429 <.0005	.283 <.0005	.362 <.0005	.344 <.0005
BP 1							1.000	.689 <.0005	-.372 <.0005	-.317 <.0005	-.026 .724
BP 2								1.000	-.263 <.0005	-.470 <.0005	-.454 <.0005
BP 3									1.000	.412 <.0005	.385 <.0005
BP 4										1.000	.583 <.0005
BP 5											1.000

SDQ 1: Emotional Symptoms, SDQ 2: Conduct Problems, SDQ 3: Hyperactivity, SDQ 4: Peer Problems, SDQ 5: Prosocial, SDQ 6: Total SDQ score, BP 1: Organisation of Experience, BP 2: Internalisation of controls, BP 3: Self-limiting features, BP 4: Undeveloped behaviour, BP 5: Unsupported development.

A look to the sub-cluster correlations with the SDQ scales shows that the sub-clusters “is emotionally secure” ($r=.344, p<.0005$), “accommodates to others” ($r=.481, p<.0005$), and “responds constructively to others” ($r=.487, p<.0005$) are most strongly related to the SDQ prosocial scale. The strongest relationship of the sub-cluster “is biddable and accepts constraints” is with the hyperactivity scale ($r=-.534, p<.0005$), that of “maintains internalised standards” with conduct problems ($r=-.369, p<.0005$), while the sub-cluster that relates the most strongly with the total SDQ scale is “maintains internalised standards” ($r=-.401, p<.0005$).

Diagnostic Profile: The cluster “self-limiting features” comprises indicators that describe “different levels of awareness and functioning but have in common the lack of a normal thrust for growth” (Bennathan and Boxall 1998, p.15). This cluster relates to the emotional symptoms scale ($r=.274, p<.0005$) and the peer problems scales ($r=.192, p=.009$) in addition to the total SDQ scale. The direction of these relationships suggests that children with more self-limiting features also display more emotional symptoms and encounter greater peer problems.

The two sub-clusters that contribute to “self-limiting features” relate in a significant way to some SDQ scales. The first sub-cluster “disengaged” most strongly relates to the peer problem scale ($r=.215, p=.004$), but “self-negating” is only significantly related with the emotional symptoms scale ($r=.354, p<.0005$). The latter is the sub-cluster that is the most strongly related to the Total SDQ scale ($r=.225, p=.002$).

The cluster “undeveloped behaviour” comprises indicators that describe a child who “has had too little help in the early years to provide him with the inner resources to relate to others and engage at an age-appropriate level” (Bennathan and Boxall 1998, p.16). This cluster relates significantly with the hyperactivity scale ($r=.404, p<.0005$) and the conduct problem scale ($r=.304, p<.0005$), as well as the peer problems ($r=.168, p=0.023$) and the total SDQ scale. Those results suggest that children with more undeveloped behaviour show more hyperactivity and conduct problem symptoms, and also more problems with peers.

Three sub-clusters contribute to this cluster. Two of them, “makes undifferentiated attachments” ($r=.174, p=.019$) and “shows inconsequential behaviour” ($r=.537, p<.0005$), relate most strongly to the hyperactivity scale. The other one, “craves attachment, reassurance” relates only with the prosocial scale ($r=.152, p=.040$). “Shows inconsequential behaviour” is the sub-cluster that relates most strongly to the Total SDQ scale ($r=.431, p<.0005$).

The last cluster, “unsupported development”, comprises indicators that describe a child who has suffered a profound lack of early nurturing care, and for whom increasingly alienated and negative behaviour appear as a protection against hurt and a way to maintain self-regard. “Their anti-social behaviour may become an increasingly well-organised, internalised pattern that brings them power and satisfaction and is thus self-perpetuating and motivating” (Bennathan and Boxall 1998, p.17). This cluster relates with the conduct problem scale ($r=.577, p<.0005$), the prosocial scale ($r=-.240, p=.001$) and the hyperactivity scale ($r=.200, p=.007$) in addition to the total SDQ scale. This means that children with a highly unsupported development show more conduct problems and hyperactivity symptoms as well as less prosocial skills.

The strongest relationships of the five sub-clusters of the “unsupported development” cluster are with the conduct problem scale, namely: “avoids/rejects attachment” ($r=.245, p=.001$), “has undeveloped/insecure sense of self” ($r=.440, p<.0005$), “shows negativism towards self” ($r=.582, p=.000$) and “wants, grabs, disregarding others” ($r=.485, p<.0005$). “Shows negativism towards self” is the sub-cluster that relates the most strongly with the total SDQ scale ($r=.328, p<.0005$).

Discussion

The results of the first discrimination analysis show that children who fall in distinct categories on the SDQ obtained scores that are statistically different on nearly all the clusters of the Boxall Profile. Thus, children who are found to be normal according to the SDQ scores obtained a profile significantly different on the BP than those assessed as abnormal on the SDQ scores. These results are a first step in establishing the predictive validity of the BP. A further study should evaluate how well the BP is able to distinguish between a low-risk and a high-risk sample.

The results obtained from the correlation analysis show with a high level of confidence that the BP and the SDQ measure similar characteristics in children and constitute strong evidence for the concurrent validity of the BP. Major differences exist between the SDQ and the BP. The main one is probably the divergence in the theoretical background of the two instruments. Effectively, the 25 items that constitute the SDQ were selected on the basis of nosological concepts that underpin the DSM-IV (APA 1994) and ICD-10 (WHO 1993) classifications of childhood psychopathology, as well as factor analysis (Goodman and Scott 1999). Conversely, items of the BP are derived from extensive direct observation of children in nurture groups and consultation with many school practitioners and a psychotherapist, as well as factor analysis.

The theoretical backgrounds also influence the structure of the questionnaires. The SDQ items are grouped to form five scales that are directly linked to the psychiatric classifications of childhood psychopathology. The scores obtained on these scales allow one to know whether the child's behaviours are within the normal range or not. Conversely, the BP items are arranged in a way that reflects how school staff interprets these features. In fact, each of the five clusters of the BP group together the features that are likely to result from the same kind of prior life experience. Two of those clusters put together features that are seen as positive in a school-age child (first section). The three other clusters put together features that are seen as problematic (second section). Neither the clusters nor the sections were given a cutting point to say if a child's behaviour was normal or not. While this characteristic is rather unusual for a behaviour rating scale, it can be easily explained.

Effectively, the BP has been constructed with the intention of "helping practitioners to understand the nature of the child's difficulties and to think constructively about the sort of help that is needed" (Bennathan and Boxall 1998, p.5). Hence, even if a normal range of results is shown for each sub-cluster, no cut-off point is proposed for determining if the child is performing within the normal or abnormal range. In fact, the authors of the BP state that "human beings and their life situations are too complex to be summed up in scores" (Bennathan and Boxall 1998, p.5). The BP is intended to refine teachers' observations, deepen their understanding of children's difficulties and inform their professional judgement. Thus, this instrument is much more likely to be used in the school context, by practitioners, than in epidemiological research. The BP might also be of value in experimental research, to measure the effect of an intervention. Conversely, the SDQ, with its well-defined cut-off points, is a very useful tool for "epidemiological research, as well as in

routine clinical and educational practice” (Goodman and Scott 1999, p.17), but is of less practical use for diagnostic purposes to teachers in their daily work with children. It is, however, a potentially useful instrument for charting behavioural change over time among groups of children.

Another major difference that is worth noting between both tools concerns their length. The SDQ examines 25 attributes, the majority of which are negative. The BP asks about 68 attributes, half of which are positive and half negative. The larger number of items and the somewhat complex way of grouping them make the BP a less straightforward tool of evaluation. However, there is anecdotal evidence to suggest that staff in schools who have used the BP believe that the time used to complete and analyse the instrument is justified and that this process contributes constructively to their classroom thinking and planning (Bennathan and Boxall 1998).

Conclusion

This study has shown that the BP and the SDQ measure comparable features in children. This research constitutes the first quantitative evidence that the BP can be a reliable instrument since its validation in 1984. Other evidence supports the context validity of the BP, based on its widespread use over many years. The statistical evidence suggests that the BP is suitable for use in a research context. These results represent a first step in this direction, but we must stress that these are preliminary findings, as they are based on a relatively small and homogenous sample. In future research, it will be important to continue the validation work by attempting to replicate these findings with larger samples, which would include more diverse populations. Another important step will be to evaluate the predictive validity of the BP by evaluating how well the questionnaire is able to distinguish between low-risk and high-risk samples.

Note

This study is part of a project funded by the Nuffield Foundation, the DFEE and the Calouste Gulbenkian Foundation, directed by Professor Paul Cooper. The Research Associates on this project are Ray Arnold and Eve Boyd. The first author would like to thank FQRSC (Fonds Québécois de la recherche sur la société et la culture) for its financial support to this project.

References

- American Psychiatric Association. 1994. *Diagnostic and statistical manual of mental disorders* (4th ed). Washington, DC: Author.
- Bennathan, M. and Boxall, M. 1996 *Effective Intervention in Primary Schools - Nurture Groups*. London: David Fulton Publishers
- Bennathan, M., and Boxall, M. 1998. *The Boxall Profile: A guide to Effective Intervention of Pupils with Emotional and Behavioural Difficulties*, East Sutton: Association of Workers for Children with

Emotional and Behavioural Difficulties.

- Bennathan, M. and Boxall, M. 2000 *Effective Intervention in Primary Schools. Nurture Groups*. London: David Fulton Publishers.
- Cooper, P.W., and Whitebread, D. 2007. The Effectiveness of Nurture Group on Student Progress: Evidence from a National Research Study. *Emotional and Behavioural Difficulties*, 12(3), 171-190.
- Goodman, R. 1997. The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.
- Goodman, R. 1999. The extended version of the strengths and difficulties questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry*, 40, 791-799.
- Goodman, R., Meltzer, H. and Bailey, V. 1998 The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry*, 7, 125-130.
- Goodman, R., and Scott, S. 1999. Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is small beautiful? *Journal of Abnormal Child Psychology*, 27, 17-24.
- World Health Organization. 1993. *The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research*. Geneva, Switzerland.