The Development and Validation of a teacher-reported Low-Level Classroom Disruption Scale (LLCD-S)

Abstract

Low-level classroom disruption (LLCD) is characterised by pupils swinging on chairs, whispering or fidgeting in class. This paper provides initial data on the development and validation of the teacher-rated Low-Level Classroom Disruption Scale (LLCD-S), with two samples of primary pupils. Exploratory factor analysis in Study 1 (*N*=120) revealed one factor accounting for 61% of the variance; supported by confirmatory factor analysis in Study 2 (*N*=274), with one factor accounting for 63% of the variance. Both studies reported high Cronbach's alpha values of .82 and .93. The evidence supports LLCD being a unidimensional construct, measured by the eight item LLCD-S. Weak convergence validity was found between the LLCD-S and the Strengths and Difficulties Questionnaire's (SDQ, Goodman, 1997) externalising behaviours: conduct problems and hyperactivity. This preliminary evidence indicates that LLCD-S is a valid and reliable measure of low-level classroom disruption. Further research is needed to test the utility of the LLCD-S across different levels of education, cultures and as a pupil-reported measure.

Keywords: low-level classroom disruption scale (LLCD-S); low-level classroom disruption, teacher-report scale; scale validation; primary school.

Word Count:7310

In a typical primary classroom an individual incident of low-level disruption, such as a pupil calling out or whispering to a peer, may seem like a relatively minor misdemeanour (Clunies-Rosset, Little, & Kienhuis, 2008). In comparison to behaviours characterised as high-level, such as bullying and aggression, the term low-level can imply such behaviours are less impactful on the classroom climate. However, low-level classroom disruption (LLCD) occurs at a high frequency in schools, and teachers consistently identify it as the number one behavioural issue across both primary and secondary schools in England (Bennett 2017: Elton 1989; Ofsted, 2014; Steer 2005). The Office for Standards in Education, Children's Services and Skills (Ofsted) estimated that LLCD has a negative effect on the education of over 700,000 pupils across the UK with pupils facing 'something of a lottery' (Ofsted, 2014, 5) of being in a classroom where teaching can take place relatively uninterrupted by incidents of LLCD. Bennet (2017) argues this prevalence rate is underestimated and that LLCD has become a toxic element in the UK classroom. These concerns are reflected in recent educational policy where LLCD has been identified as a key concern, and a major focus of a national government-funded scheme to support schools in reducing such behaviour (Department of Education 2019).

Even though teachers report confidence in handling incidences of LLCD (Ofsted, 2014), they also report greater workplace stress associated with the wearing effect of constant and repetitive interruptions to their teaching (Ofsted, 2014; Scott, Hirn, and Alter 2014; Wheldall and Merrett, 1988; Wheldall 1991). Importantly, having responsibilities for wider problems not necessarily within their individual control, such as managing LLCD, is associated with a number of negative outcomes for teachers, namely: emotional exhaustion (Blasé 1986); lower morale (Maslach, Schaufeli, and Leiter 2001); maladaptive self-efficacy (Brouwers and Tomic 2000); and losing enthusiasm and idealism for teaching, which are all factors related to professional burnout (Freudenberger 1974; Kerr and Valenti 2009). In a

meta-analysis, Aloe, Amo, and Shanahan (2014) demonstrated a link across several studies between teachers' low classroom management self-efficacy and three dimensions of burnout including emotional exhaustion, depersonalisation and lowered personal accomplishment. In line with this, Skaalvik and Skaalvik (2017) found significant correlations between teachers' perceived negativity towards interruptions caused by LLCD and their feeling of emotional exhaustion.

The impact of LLCD on pupil's learning is also significant. Emmer, Everston and Worsham, (2009) suggest that minimal distraction enables effective teaching and learning to take place, with more on task time correlated with greater learning gains. In contrast, a dysfunctional atmosphere in a classroom can negatively affects pupil attainment and academic success (Haydn et al. 2014). Longitudinal research by Duncan et al (2007) found that early disruptions to attention in class at 5–6 years old strongly predicted poor reading and maths achievement at 11–12 years old. Importantly, this result controlled for cognitive ability and was similar across gender and socioeconomic status. Furthermore, the relationship between a disruptive classroom environment, poor attainment and lower academic success has been found to extend into early adulthood; with attention problems at primary school predicting lower academic achievements at 17 years, whilst controlling for socioeconomic status and IQ (Breslau et al. 2009).

Although the negative effects of LLCD on teaching and learning are well-documented in educational reports and policy documents, empirical research specifically quantifying and reporting on LLCD is sparse. The notable absence of research could be due to the lack of a suitable tool specifically designed to measure LLCD. Elton (1989) highlighted educational concerns over the accurate recording of LLCD concluding that, 'in the absence of national statistics the problem [LLCD] itself could not be directly measured. Any estimate would have to be based mainly on teachers' perceptions' (p.59). More recently, Bennett (2017) has called

for ‘a national standardised method for capturing data on school behaviour’ (p.9) in order to record a range of behaviours including LLCD. The present study sought to address these gaps by designing a standardised scale to specifically measure low-level disruptive behaviours in the classroom. First, it is important to outline the behavioural exemplars that define LLCD, and to differentiate LLCD behaviours from other forms of classroom disruption, named here as high-level behaviours. A delineation of LLCD characteristics now follows.

Swinging on a chair, or fidgeting, can comprise a single act of LLCD which is typically low in intensity or power (Sullivan et al. 2014). LLCD has been described as presenting no physical threat or destruction to others or to school property (Kreisberg 2017) and innocuous and/or passive in nature (Beaman and Wheldall 1997). Conversely, a single act of high-level disruption (e.g. such as a kicking/ shouting at a peer or bullying) is of a high intensity and power, typically aggressive, non-compliant and extreme in nature (Wallace 2017). A single episode of high-level disruption will, as a rule, result in a high enough disturbance for teaching to be suspended, and the perpetrator excluded from the room (Hayden and Dunne 2001). These behaviours tend to involve only a single child and are relatively infrequent, which can reduce their overall impact on teaching and learning. In contrast, and fundamental to its definition, LLCD occurs at a high frequency, thus effecting classroom functioning more regularly. Although low in intensity, the rate at which incidents of LLCD occur can result in teachers having to implement a range of behaviour management strategies, which interrupts and reduces instruction time. Ofsted (2014) reported that 20% of teachers identified interruptions caused by LLCD in every lesson, accounting for up to an hour a day of lost learning time for some pupils. Moreover, due to the high frequency of LLCD and its management at a classroom level, the impact of LLCD greater and is felt across the whole class (Hall and Hayden 2007; Swinson 2010).

Qualitative accounts from teachers clearly differentiate between the characteristics of LLCD and high-level behaviours, where LLCD is described as persistent, common and an ongoing challenge (Bennett 2017; Ofsted, 2014; Wallace 2017). Estutgo-Deu and Sala-Roca (2010) found that high-level disruption was reported infrequently and was less concerning for teachers than LLCD. In their study of Spanish primary schools, they found that factors associated with LLCD (e.g. unauthorised talking) were the most frequent behavioural problem presented in class (33%), while disruptions of a highlevel (e.g. personal confrontations) were least frequent (12%). Similar evidence was found in Australian primary schools (N = 1380 teachers), with 50% of teachers stating that LLCD factors occurred several times a day, whilst 93% of the teachers reported high-level disruption not occurring at all (Sullivan et al. 2014).

The main characteristics of LLCD are therefore low intensity, passive in nature, high frequency and typically disruptive for the whole classroom. The Ofsted report entitled Low-level disruption in classrooms: below the radar (2014) presented behaviours that captured these main characteristics, as reported by a survey of teachers across primary and secondary schools in England. The radar report asked 1,048 teachers to state the most prevalent behaviours that disrupt their classroom. The top three reported were: calling out (over 50% of teachers reported), disturbing other children (almost 50% of the teachers reported) and fidgeting and fiddling with equipment (more than 33% of teachers reported), followed by talking and chatting, not getting on with work, purposely making noises to gain attention, answering back or questioning instructions and, swinging on chairs. The behaviours identified by teachers in the Ofsted report (2014) were used in the current study to underpin the construction and validation of the Low-Level Classroom Disruption Scale (LLCD-S), by forming the basis of the LLCD-S scale items.

There are many teacher/carer -rated measures readily available to assess a variety of pupil behaviours. For example, the Children's Behaviour Questionnaire (Rothbart et al. 2001) assesses temperament (extraversion/surgency, negative affectivity and effortful control) in children aged 3–7 years old and the Sutter-Eyberg Student Behaviour Inventory (Eyberg and Pincus 1999) that captures both the degree to which a behaviour is problematic and its intensity in children aged 2–16 years. The mostly widely used measure is the Strengths and Difficulties Questionnaire (SDQ: Goodman 1997), a behavioural screening tool for 3–16 year olds, which is is often used in clinical settings to assess positive and problematic behaviours across five sub-scales (emotional problems, peer problems, conduct problems, hyperactivity, and prosocial behaviour). All of these aforementioned measures quantify aspects of childhood behavioural problems; however, none specifically measure behaviours associated with the characteristics of LLCD.

An exception to this, is the recently developed Pupil Behaviour Questionnaire (PBQ: Allwood et al. 2018). This is the first scale aimed specifically at quantifying behaviours which are related to LLCD within a community sample (N= 2074, age range 4 to 9 years). This teacher-rated scale includes the following items: talking out of turn, interrupting other pupils, making unnecessary noises, making cheeky or rude remarks to the teacher, verbal abuse towards other pupils, and physical aggression towards other pupils. Although these six items achieved Cronbach's Alpha values of .70 to .90, indicating good to very good internal consistency, the two items relating to verbal abuse and physical aggression are more closely aligned with high-level behaviours and diverge from the characterisation of LLCD identified by teachers (Ofsted, 2014). Furthermore, Allwood et al (2018) compared the scale to the clinically-based Strengths and Difficulties Questionnaire (SDQ: Goodman 1997) to assess the construct validity of the PBQ. They found moderate convergence between the PBQ score and the SDQ total difficulties score (r = .59). On closer inspection, strong associations were

evident between the PBQ and conduct problems (r = .67) and hyperactivity (r = .72) which together indicate externalising behaviour. These moderate to strong convergent associations suggests that both scales may be measuring similar underlying constructs. Moderate divergence was found between the PBQ score and prosocial behaviour (r = −.53), suggesting that both scales may be measuring opposing underlying constructs. Weak to no associations were evident between the PBQ and peer problems (r = .19) and emotional symptoms (problems) (r = .01) which indicate internalising behaviour. These results suggest a strong association between externalising behaviours and LLCD as measured by the PBQ. This diverges from a definition of LLCD as presenting no physical threat or destruction to other pupils or to school property (Kreisberg 2017), and as being more passive in nature (Beaman and Wheldall 1997). Although Allwood et al (2018) concluded that the similarities with the SDQ represented good construct validity of the PBQ, we argue this is does not support the notion of LLCD as being distinct in nature and impact from high-level disruptions, in which are behaviours more closely associated with conduct problems and hyperactivity (Bennett 2017; Ofsted, 2014; Sullivan et al. 2014; Wallace 2017) and that the PBQ may not be a reliable measure of LLCD.

The present research The present paper aimed to report on the construction and validation of a scale to quantify levels of teacher-reported LLCD in primary schools. Eight items taken directly from the Ofsted report (2014), which specifically reported on LLCD, were used to construct this new LLCD-S. These were (Q1) talking and chatting, (Q2) disturbing other children, (Q3) calling out, (Q4) not getting on with work, (Q5) purposely making noises to gain attention, (Q6) fidgeting and fiddling with equipment (Q7) answering back or questioning instructions and, (Q8) swinging on chairs. As with exploratory analyses no priori hypothesis relating to these factors and patterns was predicted.

Construct analysis was also carried out. Convergent validity was assessed, correlating the LLCD-S total score with the SDQ total difficulties score and the sub-scales of externalising behaviour. We predicted weak to moderate associations with the LLCD-S, indicating convergent validity. Divergent validity was assessed comparing the LLCD-S total score with the SDQ sub-scale score of prosocial behaviour. In keeping with the view that LLCD is not overtly associated with anti-social behaviour (Kreisberg 2017), a moderate negative correlation was predicted. Additional comparisons were carried out between the LLCD-S score and the SDQ sub-scales of internalising behaviour. No direct associations were predicted between LLCD and internalising behaviour.

Given that the initial analysis of the LLCD-S was limited to the construction sample (Study 1) it was vital to test whether the scale properties would remain the same when applied to another sample. For Study 2, a cross-validation of the proposed unidimensional LLCD-S was investigated, including internal reliability and one-dimensionality, with an independent sample of primary age pupils (N= 274). It was hypothesised that LLCD as measured by the eight item LLCD-S is a one-dimensional construct.

Method

Participants

Four of the five participating schools were recruited via existing contacts of the main researcher. The fifth school was recruited via a letter drop to primary schools within a 25-mile radius of the main researcher's base in the county of Kent, UK. Both the schools for Study 1 (referred to as School 1 and School 2) and two of the schools in Study 2 (referred to as School L and School M) were located in the same urban area. The third school for Study 2 (referred to as School G) was located in a rural area, in the county of Cambridgeshire, UK. All five schools, were mixed gender and similar in cohort size (between 348 to 412 pupils) Ofsted inspections reported the schools as follows: School 1 'required improvement' (2015),

School 2 ‘satisfactory’ (2012), School G ‘good’ (2014) and School M ‘good’ (2015). No Ofsted data was available at that time for School L. Table 1 details the Office of National Statistics (2016) data, highlighting the characteristics of the geographical areas the schools were drawn from and the national figures.

[Insert Table 1 here]

The sample selection was determined by age. Adolescence (12 years +) can be regarded as a turbulent period in the behavioural trajectory of childhood (Steinberg 2005; Youniss and Smollar 1985). The onset of puberty has been associated with hormonal changes that can influence behaviour (Steinberg 2005), sometimes generating behaviour problems not previously presented (Harms et al. 2014). Therefore, it was advantageous to recruit a pre-adolescent sample, in order to limit such behavioural disturbances.

To allow for familiarity to have formed between the class teacher and their pupils, all data for both studies were collected during the final academic summer term. Study 1 took place across two primary schools located in the county of Kent, UK. LLCD data was collected from the class teachers (N= 4) reporting on the pupil sample (N= 120). The pupils self-reported their year group, age and gender. The pupil sample was spread across two year groups (5 and 6) with an age range of 9–11 years old (M_{age} = 10.29, SD = .64). Of the total pupil sample, 49.2% were in year 5 and 50.8% were in year 6, 59% of pupils identified as male and 41% identified as female. Study 2 took place across three primary schools located in the counties of Kent and Cambridgeshire, UK. Using the LLCD-S, data was collected from the teachers (N= 8) reporting on the pupil sample (N= 248). The pupils self-reported their year group, age and gender. The pupil sample was spread across two year groups (year 4 and 5) with an age range of 8–10 years (M_{age} = 9.34, SD = .66). Of the total pupil sample, 47.8%

were in year 4 and 52.2% were in year 5, 49% of pupils identified as male and 51% identified as female.

Measures

Low-level classroom disruption. The intent of this scale construction was to generate items to measure low-level classroom disruption, specifically. Past literature has defined LLCD as having the characteristics of low intensity, high frequency and as having impact across the classroom as a whole unit (Kreisberg 2017; Sullivan et al. 2014; Swinson 2010). To reflect this definition, and to ensure a sufficient breadth of LLCD content was included, the eight highest teacher-rated behavioural issues as highlighted in the Ofsted report (2014) were selected as items. For the present study, teachers were instructed to rate how often each individual pupil in their class carried out the following eight acts: 1) talking and chatting, 2) disturbing other children, 3) calling out, 4) not getting on with work, 5) purposely making noises to gain attention, 6) fidgeting and fiddling with equipment, 7) answering back or questioning instructions and 8) swinging on chairs. Responses were rated on a three-point scale of 1 (never), 2 (sometimes), 3 (a lot), with a higher score indicating a higher presentation of LLCD. Following guidance of prior scale development work (Clark and Watson 1995), all eight items were positively worded to avoid ambiguity in the interpretation of meaning. Limiting the scale to eight items also enables the teachers to complete quick assessment on every child in the class (Slade, Thornicroft, and Glover 1999). Cronbach's alphas values were very good to excellent, with Study 1 equal to .82 and Study 2 equal to .93.

Behaviour. For Study 1, teachers also completed The Strengths and Difficulties Questionnaire (SDQ: Goodman 1997) for each pupil. The SDQ is a well-validated behavioural screening questionnaire for 3–16 year olds, typically completed for clinical diagnostic purposes. The SDQ measures emotional and behavioural changes. Consisting of five subscales (emotional problems, peer problems, conduct problems, hyperactivity, and pro-

social behaviour), there are 25 items rated on a three-point Likert scale (Not True, Somewhat True and Certainly True). A total difficulties score is derived by summing emotional problems, peer problems, conduct problems and hyperactivity. Regarded as difficulties, a higher score on the total SDQ (or on the subscales of emotional problems, peer problems, conduct problems and hyperactivity), indicates higher difficulty in establishing adaptive behaviour. For prosocial behaviour the reverse is true, with a higher score indicating a higher level of this adaptive behaviour, which is regarded as a strength. An externalising behaviour score can be created by summing hyperactivity and conduct problems, and an internalising behaviour score can be created by summing emotional problems and peer problem scores. The SDQ has a test-retest correlation of .85 (Goodman and Scott 1999) displaying strong evidence of construct validity. See Table 2 for the present study Cronbach's alpha values for the SDQ totals and subscale.

[Insert Table 2 here]

Procedure

For both studies, the head teachers granted consent for the research to take place. Each school distributed information letters regarding the study to the pupils' parents/guardians, with the option to withdraw their child/s from the study. There were minimal opt-out requests returned to the schools, Study 1 = 3, and Study 2 = 14. The class teachers (Study 1 *N*= 4, Study 2 *N*= 8) received information regarding the purpose of the study and informed that any data collected would be confidential in nature and anonymised. All the teachers approached granted their consent to participate. For Study 1, teachers received the Low-Level Classroom Disruption Scale (LLCD-S) and the Strengths and Difficulties Questionnaire (Goodman 1997). For Study 2, teachers received the Low-Level Classroom Disruption Scale (LLCD-S). For both studies, teachers received a master list of pupils' names and pupils' personal codes (Study 1 *N*= 120, Study 2 *N*= 274). Teachers completed the questionnaires in their own time,

recording each pupil's personal code along with a personal code of their own. On completion of the data collection, all participants (parents, pupils and teachers) received debriefing forms. These contained full details of the study, ethical issues such as post hoc withdrawal from the study and information about help/support lines should they require this.

Results: Study 1

Descriptive statistics

On average, pupils in the sample were 10.29 years old (*SD* = .64; range 9–11) at the time of the data collection. The sample was made up of 59 boys and 61 girls (*N*= 120). Table 3 presents the means, standard deviations and bivariate correlations of the LLCD-S items from Study 1. All items of the LLCD-S were positively correlated and larger than .3.

[Insert Table 3 here]

Exploratory factor analysis (EFA)

EFA assessed the eight items of the teacher-rated LLCD-S with maximum likelihood estimator, using SPSS (IBM). The sample size and the strength of relationship between the items indicated suitability of the data for EFA. With a Kaiser-Meyer-Olkin value for the data set of .8, being greater than the recommended .5 (Kaiser 1970). The strength of the relationship between the items considered Pearson's correlations and revealed the presence of all coefficients larger than .3. Bartlett's test of sphericity (Bartlett 1954) tested the overall multivariate correlations within the correlation matrix and was significant ($\chi 2$ (28) = 542.64, $p < .001$). Thus, indicating normality of distribution, supporting the factorability of the data (Table 3). Following the eigenvalue rule, stating only eigenvalues larger than one retained (Howitt and Cramer 2017), the EFA analysis identified the existence of one factor. With an eigenvalue equal to 4.88 achieving a total variance in the data of 60.92%. As Table 4

indicates, the component matrixes revealed very strong loadings on this one factor for all eight items of the measure (>.50).

[Insert Table 4 here]

Inspection of the scree plot revealed a clear elbow with one point above this, supporting a one-factor solution (Figure 1). Parallel Analysis further supported these results, which showed only one component with eigenvalues exceeding the corresponding criterion values for a randomly generated data matrix of the same size (8 variables x 120 respondents). These results demonstrate that all eight items converge on the same factor, indicating one salient construct underling the LLCD-S item scores. Rotation did not take place, as all eight items loaded sufficiently onto one factor. The LLCD-S demonstrated excellent internal consistency with a Cronbach's alpha value of .82. This result compared very favourably with the recommended value for scales used in research of above .6 (Nunnally 1978).

[Insert Figure 1 here]

Convergent Validity

Convergent validity was investigated by calculating Spearman's correlations coefficients between the LLCD-S and the SDQ total scale, and sub-scales of the SDQ (Goodman 1997). As was predicted small positive correlations were found between the LLCD-S score and the SDQ total difficulties and the hyperactivity sub-scale scores, indicting weak similarities. These similarities were noticeably weaker than the moderate similarities that were found for these convergent correlations by Allwood et al (2018). Contrary to the prediction, a medium positive correlation was found between the LLCD-S total score and the conduct problems score, as measured by the SDQ sub-scale. This correlation was similar to the correlation found between LLCD, as measured by the PBQ, and conduct problems as measured by the

SDQ sub-scale during previous research (Allwood et al. 2018). As was predicted convergent investigations carried out between the LLCD-S score and the SDQ externalising behaviour scale score found a medium correlation, indicting a moderate similarity (Table 5).

Divergent Validity

To assess divergent validity Spearman's correlation coefficients were computed. In keeping with the prediction a medium negative correlation was found between the LLCD-S total score and the SDQ prosocial behaviour sub-scale score. This correlation value was similar to that previously reported between the PBQ and the SDQ prosocial behaviour sub-scale during previous research (Allwood et al. 2018). Contrary to the prediction stating that no association would be found, a significant correlation was found between the LLCD-S score and the SDQ sub-scale score of peer problems, however this was a weak association. Once again this was similar to the correlation found by previous research (Allwood et al. 2018) between the PBQ and the SDQ sub-scale of peer problems. As was predicted, no significant associations were found between the LLCD-S and the SDQ total internalising behaviour sub-scale or the sub-scale of emotional problems (Table 5).

[Insert Table 5 here]

Results: Study 2

Descriptive Statistics

On average, pupils in the sample were 9.34 years old (*SD* = .66; range 8–10) at the time of the data collection. The sample was made up of 121 boys and 127 girls (*N*= 248). Table 6 presents the means, standard deviations and bivariate correlations of the LLCD-S items from Study 2. All correlation coefficients between the items of the LLCD-S were greater than .3.

[Insert Table 6 here]

Confirmatory Factor Analysis (CFA)

Considering the one-factor solution identified in Study 1, CFA was conducted to test the following hypothesis: Low-level classroom disruption, as measured by the eight item LLCD-S, is a one-dimensional construct. Replicating the model from Study 1, the CFA model for the teacher-reported LLCD-S constrained all eight items to load onto one factor. Model fit assessed the CFA indices, indicating a good fit: X^2 (272) = 174.33, p< .001, SRMR = .052, CFI = .90 and TLI = .86 (Hu and Bentler 1999). Moreover, all standardized factor loadings were statistically significant, ranging from .64 to .88 (Table 7). Overall, CFA results indicate adequate factor structure for the cross validation sample. Reflecting Study 1, the Cronbach's reliability coefficients for Study 2 recorded an alpha value of .93 indicting excellent internal consistency.

[Insert Table 7 here]

Discussion

These studies describe the construction and factor structure of the teacher-reported Low-Level Classroom Disruption Scale (LLCD-S), providing preliminary evidence of the reliability and validity of one factor. First, the study presented previous literature and outlined differences between the concept of low-level classroom disruption and high-level classroom behaviours. LLCD has been consistently defined by teachers as being of low intensity, passive in nature, high frequency and typically disruptive for the whole classroom (Bennett 2017; Ofsted, 2014; Wallace 2017). Whereas, high-level classroom behaviours are conversely characterised by their high intensity, low frequency and typically disruptive for only the perpetrator of the maladaptive behaviour (Bennett 2017; Ofsted, 2014; Sullivan et al. 2014; Wallace 2017). Based on this differential and evidence reported in the Low-level disruption in classrooms: below the radar report (Ofsted, 2014), which specifically investigated LLCD, the eight items capturing the behaviour exemplars of LLCD were defined

as: (Q1) talking and chatting, (Q2) disturbing other children, (Q3) calling out, (Q4) not getting on with work, (Q5) purposely making noises to gain attention, (Q6) fidgeting and fiddling with equipment, (Q7) answering back or questioning instructions, (Q8) swinging on chairs. These eight behaviour exemplars were included as the items for the development of the LLCD-S.

Due to the exploratory nature of this new scale no priori hypothesis was forecast. The Study 1 values for the KMO and the Bartlett's sphericity test revealed that the sample of 120 was large enough for exploratory factor analysis (EFA) to take place and that scores were normally distributed. EFA yielded a one-factor structure, with all eight items loading significantly onto this one factor, explaining 61% of the total variance. As for any scale development it is imperative that the internal consistency of the scale properties is tested on additional data sets; therefore, Study 2 evaluated the LLCD-S with a new sample of 274 primary pupils. The hypothesis that low-level classroom disruption (as measured by the eight item LLCD-S) is a one-dimensional construct was upheld. Confirmatory factor analysis (CFA), based on the previous EFA results from Study 1, supported a single factor model and explained 63% of the total variance. Estimates of the internal consistency of a scale should range from .7 to .9 to indicate reliability (McCrae et al. 2011). Encouragingly, the single factor scale showed strong internal consistency for both Study 1 and Study 2 (.82 and .93 respectively), indicating that all eight items were measuring the same underlying concept of LLCD and that the LLCD-S is a highly reliable scale across two different samples of primary aged pupils. These excellent internal consistency results allow for the preliminary conclusion that the LLCD-S is an accurate measure of the presentation of low-level classroom disruption with a primary school sample. It therefore provides education practitioners with a much needed and long awaited means of systematically capturing LLCD (Bennett 2017; Elton 1989).

Adding strength to the development of this new measure, Study 1 assessed the construct validity of the LLCD-S by concurrently collecting teacher-rated scores from the SDQ (Goodman 1997). As expected, results showed strong divergent validity between the LLCD-S and prosocial behaviour as measured by the SDQ; whereby moderate convergent validity was found between LLCD-S and externalising behaviour, and weak convergent validity between LLCD-S and internalising behaviour. The LLCD-S was designed to capture low-level behaviours in classroom settings while the SDQ was designed to capture higher level behaviours in clinical populations, and therefore while we expected some convergence, the two scales are measuring distinct underlying constructs. In comparison, the convergence between SDQ externalising behaviours and the LLCD-S was notably lower in strength than convergence between SDQ externalising behaviours and the Pupil Behaviour Scale (Allwood et al. 2018).

Crucially, LLCD has been consistently highlighted as the number one behavioural issue in primary schools across the UK; causing ongoing and greater concerns compared to high-level behaviours (Bennett 2017; Estutgo-Deu & Sala-Roca, 2010; Ofsted, 2014; Wallace 2017). Therefore, it is imperative that when aiming to measuring LLCD behaviour in the classroom, only behaviours specifically conducive to low-level disruption are captured. The timing of this present study is of added importance in light of the English Government's recent announcement that in September 2020 a programme of research will be launched to tackle bad behaviour in the classroom, specifically including investigations of LLCD (Department of Education 2019). The LLCD-S could provide an important evidenced-based tool with which to measure LLCD and evaluate interventions.

A major strength of this report is the replication of the scale reliability over two studies. Despite this, there are some limitations to consider. First, the present study aimed to limit the capturing of behavioural disturbances associated with adolescence, therefore both

samples were restricted to pre-adolescent primary pupils aged between 8–11 years; however, this limits the generalisability of the results. As LLCD is also reported as a significant issue at secondary education, it is highly recommended that future research should look to expand the sample age range to include adolescence and/or post-adolescence. This would enable important investigations to observe the influences on LLCD, and the changes to LLCD both during and across key developmental stages. Second, the research locations were limited to the counties of Kent and Cambridgeshire, UK. Future investigations of LLCD should look to widen the research areas to include school samples from across the UK, and beyond; in order to capture a more diverse sample, which would also enable the exploration of socio-economic factors. Third, this paper only reports on the observer-rated scores of the LLCD-S from class teachers. Future studies could evaluate a pupils' self-reported LLCD-S in order to reduce the risk of common method variance. Finally, the cross-sectional nature of this study does not allow for test-retest assessment of the external reliability of the LLCD-S. Having the same sample report levels of LLCD, over two or more separate data collection waves, would allow correlations between the time points to be calculated. Therefore, future studies could overcome this limitation by implementing a longitudinal design.

Concluding Remarks

In conclusion, LLCD has been consistently emphasised as the number one behavioural issue in primary schools, with a negative impact on both teachers and pupils (Bennett 2017: Elton 1989; Ofsted, 2014; Steer 2005). Considering this, and addressing a recognised gap in the literature, the LLCD-S can be effectively utilised for screening and/or as an outcome measure to accurately record low-level classroom disruption presentation at primary school level (Elton 1989; Ofsted, 2014). Importantly, the LLCD-S focuses specifically on LLCD clearly addressing only low-level maladaptive behaviours, which differentiates the LLCD-S from prior measures that include high-level maladaptive

behaviours. This scale would be beneficial to quantify levels of LLCD across individual pupils, classrooms and schools. Further exploration of the scale is required across time with various age populations, across the UK and beyond, and additionally as a pupil self-report measure. Disclosure statement No potential conflict of interest was reported by the authors.

References

Allwood, M., K. Allen, A. Price, R. Hayes, V. Edwards, S. Ball, O. C. Ukoumunne, and T. Ford. 2018. "The Reliability and Validity of the Pupil Behaviour Questionnaire: A Child Classroom Behaviour Assessment Tool." Emotional and Behavioural Difficulties 23 (4): 361–371. doi:10.1080/13632752.2018.1478945.

Aloe, A. M., L. C. Amo, and M. E. Shanahan. 2014. "Classroom Management, Self-efficacy and Burnout: A Multivariate Meta-analysis." Educational Psychology Review 26 (1): 101–126. doi:10.1007/s10648-013-9244-0.

Bartlett, M. 1954. "Significance Test for Sphericity of a Normal N-variate Distribution." Journal of the Royal Statistical Society 16: 296–298.

Beaman, R., and K. Wheldall. 1997. "Teacher Perception of Troublesome Classroom Behaviour: A Review of Recent Research." Special Education Perspectives 6: 49–56.

Bennett, T. 2017. Creating a Culture: How School Leaders Can Optimise Behaviour. UK Department for Education. https:// www.gov.uk/government/publications/behaviour-in-schools

Blasé, J. J. 1986. "A Qualitative Analuysis of Sources of Teacher Stress: Consequences for Performance." American Educational Research Journal 23 (1): 13–40. doi:10.3102/00028312923001013.

Breslau, J., E. Miller, N. Breslau, K. Bohnert, V. Lucia, and J. Schweitzer. 2009. "The Impact of Early Behaviour Disturbances on Academic Achievement in High School." Paediatrics 123 (6): 1472–1476. doi:10.1542/peds2008-1406.

Brouwers, A., and W. Tomic. 2000. "Disruptive Student Behaviour." Perceived Self-Efficacy and Teacher Burnout. http:// files.eric.ed.gov/fulltext/ED450120.pdf

Clark, L. A., and D. Watson. 1995. “Constructing Validity: Basic Issues in Objective Scale Development.” Psychological Assessment 7 (3): 309. doi:10.1037/1040-3590.7.3.309.

Clunies-Ross, P., E. Little, and M. Kienhuis. 2008. “Self-reported and Actual Use of Proactive and Reactive Classroom Management Strategies and Their Relationship with Teacher Stress and Student Behaviour.” Educational Psychology 28 (6): 694–710. doi:10.1080/01443410802206700.

Department of Education. 2019. “School Backed to Tackle Bad Behaviour.” May. https://www.gov.uk/government/news/ schools-backed-to-tackle-bad-behaviour

Duncan, G. J., C. J. Dowsett, A. Claessens, K. Magnuson, A. C. Huston, P. Klebanov, . . . H. Sexton. 2007. “School Readiness and Later Achievement.” Developmental Psychology 43 (6): 1428. doi:10.1037/0012-1649.43.6.1428.

Elton, B. R. 1989. Discipline in Schools: report of the Committee of Enquiry chaired by Lord Elton. HMSO. http://www. educationengland.org.uk/documents/elton/elton1989.html

Emmer, E. T., C. M. Evertson, and M. E. Worsham. 2009. Classroom Management for Middle and High School Teachers. 6th ed. Boston, MA: Allyn & Bacon.

Esturgo-Deu, M. E., and J. Sala-Roca. 2010. “Disruptive Behaviour of Students in Primary Education and Emotional Intelligence.” Teaching and Teacher Education 26 (4): 830–837. doi:10.1016/j.tate.2009.10.020.

Eyberg, S. M., and D. Pincus. 1999. Eyberg Child Behaviour Inventory Professional Manual. Odessa: Psychological Assessment Resource, Ins. Freudenberger, H. J. 1974. “Staff Burn-out.” Journal of Social Issues, 30(1), 159–165. doi:10.1111/j.1540-4560.1974. tb00706.x.

Goodman, R. 1997. “The Strengths and Difficulties Questionnaire: A Research Note.” Journal of Child Psychology and Psychiatry 38 (5): 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x.

Goodman, R., and S. Scott. 1999. “Comparing the Strengths and Difficulties Questionnaire and the Child Behaviour Checklist: Is Small Beautiful?” Journal of Abnormal Child Psychology 27 (1): 17–24. doi:10.1023/A:1022658222914.

Hall, N., and C. Hayden. 2007. “Is ‘Hate Crime’ a Relevant and Useful Way of Conceptualising Some Forms of School Bullying.” International Journal on Violence and Schools 3: 3–24.

Harms, M. B., V. Zayas, A. N. Meltzoff, and S. M. Carlson. 2014. “Stability of Executive Function and Predictions to Adaptive Behaviour from Middle Childhood to Pre-Adolescence.” Frontiers in Psychology 5: 331. doi:10.3389/fpsyg.2014.00331.

Hayden, C., and S. Dunne. 2001. Outside, Looking In: Children’s and Families’ Experiences of Exclusion from School. London: Children’s Society

Haydn, T., A. Stephen, J. Arthur, and M. Hunt. 2014. Learning to Teach History in the Secondary School: A Companion to School Experience. London and New York: Routledge.

Howitt, D., and D. Cramer. 2017. Understanding Statistics in Psychology with SPSS. Essex, England: Pearson Higher Ed.

Hu, L. T., and P. M. Bentler. 1999. “Cut-off Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives.” Structural Equation Modelling: A Multidisciplinary Journal 6 (1): 1–55. doi:10.1080/10705519909540118.

Kaiser, H. 1970. “A Second Generation Little Jiffy.” Psychmetrika 35 (4): 401–415. doi:10.1007/BF02291817.

Kerr, M. M., and M. W. Valenti. 2009. “Control from Withn the Classroom: Crises or Conversations.” Reclaiming Children and Youth, 17(4), 30–34. www.researchgate.net/profile.Mary_Kerr2/publication/234674640_Controls_from_within_ the_Classroom_Crises_or_Conversations/links/57acf9a008ae3765c3baffba.pdf

Kreisberg, H. L. 2017. “Making sense of low-level behaviours in the elementary classroom: An interpretative phenomenological analysis.” Doctoral dissertation. Northeastern University.

Maslach, C., W. B. Schaufeli, and M. P. Leiter. 2001. “Job Burnout.” Annual Review of Psychology 52 (1): 397–422. doi:10.1146/aantev.psych.52.1.397.

McCrae, R. R., J. E. Kurtz, S. Yamagata, and A. Terracciano. 2011. “Internal Consistency, Retest Reliability, and Their Implications for Personality Scale Validity.” Personality and Social Psychology Review 15 (1): 28–50. doi:10.1177/ 1088868310366253.

Nunnally, J. C. 1978. “Psychometric Theory.” McGraw-Hill Book Company, Pp 86-113: 190–255.

Ofsted. 2014. “Below the Radar: Low Level Disruption in the Country’s Classrooms.” http://www.gov.uk/governemtn/ publications/below-the-radar-low-level-disruption-in-the-country’s-classrooms

Rothbart, M. K., S. A. Ahadi, K. L. Hershey, and P. Fisher. 2001. “Investigations of Temperament at Three to Seven Years: The Children’s Behaviour Questionnaire.” Child Development 72 (5): 394–408. doi:10.1111/1467-8624.00355.

Scott, T. M., R. G. Hirn, and P. J. Alter. 2014. “Teacher Instruction as a Predictor for Student Engagement and Disruptive Behaviours.” Preventing School Failure: Alternative Education for Children and Youth 58 (4): 193–200. doi:10.1080/ 1045988X.2013.787588.

Skaalvik, E. M., and S. Skaalvik. 2017. “Still Motivated to Teach? A Study of School Context Variables, Stress and Job Satisfaction among Teachers in Senior High School.” Social Psychology of Education 20 (1): 15–37. doi:10.1007/ s11218-016-9363-9.

Slade, M., G. Thornicroft, and G. Glover. 1999. “The Feasibility of Routine Outcome Measures in Mental Health.” Social Psychiatry and Psychiatric Epidemiology 34 (5): 243–249. doi:10.1007/s001270050139.

Steer, A. 2005. Learning behaviour: The report of the practitioners’ group on school behaviour and discipline. Nottingham: DfES. www.educationengland.org.uk/documents/pdfs/2005-steer-report-learning-behaviuor.pdf

Steinberg, L. 2005. “Cognitive and Affective Development in Adolescence.” Trends in Cognitive Science 9 (2): 69–74. doi:10.1016/j.tics.2004.12.005.

Sullivan, A. M., B. Johnson, L. Owens, and R. Conway. 2014. “Punish Them or Engage Them? Teachers’ Views of Unproductive Student Behaviours in the Classroom.” Australian Journal of Teacher Education 39 (6): 43. doi:10.14221/ajte.2014v39n6.6.

Swinson, J. 2010. “Working with a Secondary School to Improve Social Relationships, Pupil Behaviour, Motivation and Learning.” Pastoral Care in Education 28 (3): 181–194. doi:10.1080/02643944.2010.504221.

Wallace, S. 2017. Behaviours Management: Getting It Right in a Week. Critical Publishing, ST Albans, Herts, UK.

Wheldall, K. 1991. “Managing Troublesome Classroom Behaviour in Regular Schools: A Positive Teaching Perspective.” International Journal of Disability, Development and Education 38 (2): 99–116. doi:10.1080/0156655910380202.

Wheldall, K., and F. Merrett. 1988. “Which Classroom Behaviours Do Primary School Teachers Say They Find the Most Troublesome?” Educational Review 40 (1): 13–27. doi:10.1080/0013191880400102.

Youniss, J., and J. Smollar. 1985. Adolescent Relations with Mothers, Fathers and Friends. Chicago: University of Chicago Press.

Appendix

The Low-level classroom disruption scale (LLCD-S)

This is a chance to find out about the pupils' behaviour in your classroom. Please write your own personal code and each pupil's individual code in the spaces below. Be sure that your answers show accurately how each individual pupil behaves. Please read each item then place a tick in the box to indicate the individual pupil's level of presentation for each behaviour. Please do not talk to anyone about your answers. We will keep your answers private and not show them to anyone.

Teacher Code:	Pupil Code:		
	Tick one box only for each behaviour		
Item	Never	Sometimes	Often
Talking and chatting			
Disturbing other children			
Calling out			
Not getting on with work			
Purposely making noise to gain attention			
Fidgeting or fiddling with equipment			
Answering back or questioning instructions			
Swinging on chairs			

Scoring the LLCD-S. Each item is scored as 1-Never, 2-Sometimes, 3-Often. The scores are summed with a possible range of 8 – 24. A higher score indicates a higher presentation of LLCD.

*Table **1***

Regional Characteristics of Sample Schools

	Study 1	Study 2		National
School	1 & 2	G	L & M	
Population of Region	276,000	100,000	276,000	
Education: NVQ4 or above	25%	28%	25%	37%
Education: No Qualifications	8.3%	10.1%	8.3%	8.6%
Employed	72%	72%	72%	74%
Gross weekly full-time wage	£566.10	£504.00	£566.10	£541.00
Unemployment Level	6.8%	3.7%	6.8%	5.1%
Unemployed: Long term sick	18%	43%	18%	22.5%
Unemployed: Lone Parent	1.3%	1.1%	1.3%	1%

Table 2.

Study 1. Cronbach's alpha values for the Strengths and Difficulties Questionnaire scale and subscales (N=120).

SDQ Variable	Cronbach's alpha
Total difficulties	0.54
Conduct problems	0.76
Hyperactivity	0.87
Total externalising behaviour	0.89
Peer pressure	0.66
Emotional problems	0.86
Total internalising behaviour	0.64
Prosocial behaviour	0.85

Table 3.

Correlation matrix, means, standard deviations for Low-level classroom disruption scale (LLCD-S) for Study 1(N=120).

Item	2	3	4	5	6	7	8	*M*	*SD*
1 Talking and chatting	.62	.57	.42	.42	.59	.49	.41	1.92	.61
2 Disturbing other children	-	.69	.70	.59	.72	.61	.38	1.55	.64
3 Calling out		-	.48	.60	.58	.72	.41	1.35	.61
4 Not getting on with work			-	.42	.62	.52	.44	1.61	.65
5 Purposely making noise to gain attention				-	.60	.62	.42	1.15	.40
6 Fidgeting or fiddling with equipment					-	.66	.57	1.41	.65
7 Answering back or questioning instructions						-	.49	1.27	.54
8 Swinging on chair							-	1.22	.50

Table 4.

Exploratory factor analysis: Factor loadings for teacher-rated LLCD-S for Study 1 (N=120).

Factor Loading	Item Number	Item
.847	2	Disturbing other children
.841	6	Fidgeting or fiddling with equipment
.798	3	Calling out
.798	7	Answering back or questioning instructions
.707	5	Purposely making noise to gain attention
.703	4	Not getting on with work
.676	1	Talking and chatting
.571	8	Swinging on chair

Notes. Eigenvalue = 4.874, Percent of variance = 61%; Extraction Method: Maximum Likelihood; 1 component extracted.

Table 5.

Spearman's coefficient correlations(r_s) between the teacher completed LLCD-S and the teacher completed SDQ (totals and subscales) for Study 1 (N=120).

	r_s	*p*-value
Total LLCD-S vs total difficulties	.22	<0.05
Total LLCD-S vs conduct problems	.56	<0.01
Total LLCD-S vs hyperactivity	.34	<0.01
Total LLCD-S vs total externalising behaviour	.51	<0.01
Total LLCD-S vs peer problems	-.20	<0.05
Total LLCD-S vs emotional problems	-.09	N/S
Total LLCD-S vs total internalising behaviour	-.15	N/S
Total LLCD-S vs prosocial behaviour	-.60	<0.01

Note: N/S = no significant result

Table 6.

Low-level classroom disruption scale (LLCD-S): Bivariate correlations, means, standard deviations for Study 2 (N=274).

Item	2	3	4	5	6	7	8	*M*	*SD*
1. Talking and chatting	.70	.63	.54	.44	.51	.50	.46	1.97	.63
2. Disturbing other children		.73	.69	.61	.69	.63	.58	1.61	.69
3. Calling out			.59	.58	.62	.65	.62	1.44	.68
4. Not getting on with work				.66	.72	.62	.61	1.51	.65
5. Purposely making noise to gain attention					.78	.72	.76	1.21	.55
6. Fidgeting and fiddling with equipment						.69	.75	1.37	.65
7. Answering back or questioning instructions							.64	1.27	.56
8. Swinging on chair								1.25	.54

Table 7.

Confirmatory factor analysis: Factor loadings for teacher-rated LLCD-S in Study 2 (N=274).

Factor Loading	Item Number	Item
.877	6	Fidgeting or fiddling with equipment
.842	5	Purposely making noise to gain attention
.807	8	Swinging on chair
.803	2	Disturbing other children
.801	7	Answering back or questioning instructions
.798	4	Not getting on with work
.769	3	Calling out
.641	1	Talking and chatting

Notes. Percent of variance = 63%; Extraction Method: Maximum Likelihood.

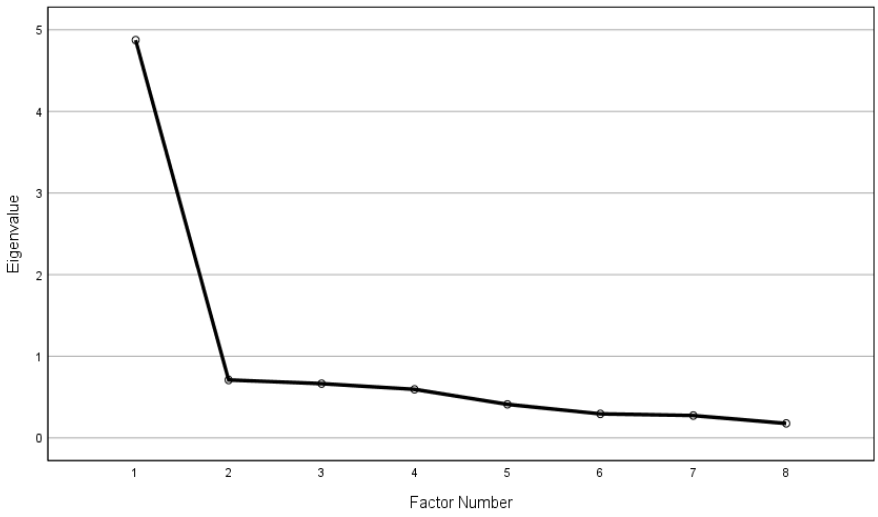

Figure 1: Scree Plot for low-level classroom disruption scale for Study 1 (*N* = 120).