A Mokken scale to assess secondary pupils' experience of violence in terms of severity

Citation for published version (APA):

Mooij, A. J. (2012). A Mokken scale to assess secondary pupils' experience of violence in terms of severity. *Journal of Psychoeducational Assessment*, 30(5), 496-508. <https://doi.org/10.1177/0734282912439387>

DOI:

[10.1177/0734282912439387](https://doi.org/10.1177/0734282912439387)

Document status and date:

Published: 01/10/2012

Document Version:

Peer reviewed version

Document license:

CC BY-NC-SA

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

<https://www.ou.nl/taverne-agreement>

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 26 Apr. 2023

Open Universiteit
www.ou.nl

Running head: MOKKEN SCALE TO ASSESS EXPERIENCE OF VIOLENCE

A Mokken Scale to Assess
Secondary Pupils' Experience of Violence
in Terms of Severity

Ton Mooij

Radboud University Nijmegen (ITS)

and Open University of the Netherlands (CELSTEC)

A Mokken Scale to Assess Secondary Pupils' Experience of Violence in Terms of Severity

Abstract

Violence assessment can potentially be improved by Item Response Theory, i.e. ordinal Mokken Scale Analysis. The research question is: Does Mokken Scale Analysis of secondary pupils' experience of violence result in a homogeneous, reliable, and valid unidimensional scale that fits all the requirements of Mokken scaling? The method used is secondary analysis of Dutch national data collected from secondary school pupils in 2008 by means of a digital school safety survey. The first random sample ($n_1=14,388$) was used to develop the scale; the second sample ($n_2=14,350$) is meant to cross-validate the first results. Pupils' experience of violence is assessed by 29 items reflecting six types of antisocial or aggressive behaviour. A Mokken scale of 25 items meets the requirements of monotone homogeneity and double monotonicity. Ordering is invariant between: Boys and girls; being born in the Netherlands or not; and feeling at home in the Netherlands or not. These results are cross-validated in sample 2. The latent construct concerns pupils' experience of violence in terms of severity, varying from verbal and mild physical violence (relatively most frequent), to combinations of social, material and severe physical violence, to very severe and serious sexual violence (relatively least frequent). Some limitations and further developments are discussed.

Keywords: Mokken scale; Item Response Theory; severity of violence; digital assessment; secondary school students

One of the main reasons researchers investigate experiences of violent behaviour is that such behaviour has various negative consequences for the person or group being victimised.

Violence-related negative effects may be expressed in feelings of anxiety, depression, low self-esteem, and loneliness (Duke, Pettingell, McMorris, & Borowsky, 2010). An experience of violence may also involve physical pain and injury or material damage, for example vandalism (Bayh, 1975). Moreover, frequent victimisation by peers at school is associated with poor academic performance (Schwartz, Gorman, Nakamoto, & Toblin, 2005). Violence perpetrated in and around school, including on the Internet, is interpreted as a threat to safety and social cohesion (Chen & Astor, 2011; Finkelhor, Omrod, Turner, & Hamby, 2011; Siu, 2011). Violence-related experiences in and around schools are therefore a major concern in the educational policy of countries such as the United States (Mayer & Furlong, 2010), Canada (Beauvais & Jenson, 2002), New Zealand (Office of the Children's Commissioner, 2010), and the Netherlands (Ministry of Education, Culture and Science, 2011).

Researchers assess violent behaviour or pupils' experience of violence within the context of different behaviour areas and by means of different procedures. Antisocial behaviour and experiences can be measured by means of a self-reporting instrument such as the Olweus Bully/Victim Questionnaire (Solberg & Olweus, 2003). Wang, Iannotti, Luk, and Nansel (2010) have extended this instrument to examine the co-occurrence of victimisation in subtypes of bullying, for example physical and verbal bullying, social exclusion, spreading rumours, and cyber bullying. Mynard and Joseph (2000) have developed a multidimensional psychometric scale describing different types of peer victimisation. Nylund, Bellmore, Nishina, and Graham (2007) have classified secondary school pupils by their victimisation experiences and shown that victim groups are best understood according to level of severity rather than type. Felix, Furlong, and Austin (2009) clustered pupils with similar characteristics, resulting in five victimisation subgroups. Mooij (2011a) has focused on

relationship patterns between personal and other characteristics of secondary school pupils and their motives as a victim, perpetrator, or witness of six types of violent behaviour, in relation to the complementary social roles of other pupils, teachers, other school staff, and pupils' relatives. In the same vein, Mooij (2011b) has examined social interaction patterns between the personal and school characteristics of secondary school teachers and their experience of violence in differentiated ways.

Because violence is measured in different areas, in different ways, and within different groups, and because the results are used in different designs, it is somewhat difficult to compare the research findings of multiple studies or to assess their relevance for identifying or reducing the negative effects described above. Michie and Cooke (2006) have summarised the problems involved in violence measurement as: Multidimensionality in assessment; non-empirical ordering of violent acts; inclusion of indiscriminating items; and differential precision of measurement across the range of seriousness. They themselves used the MacArthur Community Violence Screening Instrument (MCVSI) to interview 250 male prisoners between 18 and 40 years of age. They applied Item Response Theory (IRT) and found that the instrument's items were not ordered correctly in terms of the severity of the underlying trait; additional items were required to improve discrimination.

Given the diversity of violence concepts and assessment procedures, it is worth looking more closely at IRT. IRT models generally show how the items in a scale function relative to one another and where each item is situated on the continuum of an underlying construct from low to high severity or difficulty (cf. Schafer, 1996). When the items form a unidimensional scale, it is appropriate to combine them in order to reach a total score that can be related to the scores of other variables. IRT may therefore solve some of the problems associated with the measurement of violence. An example is given by Regan, Bartholomew, Kwong, Trinke, and Henderson (2006), who have evaluated the structure of the physical

violence scale, part of the Conflict Tactics Scales (CTS). These researchers have determined the ordering of items used to assess 14 acts of physical violence within heterosexual relationships in a sample of women and men.

There has been little IRT-based research on violent behaviour and the associated experiences (Regan et al., 2006). Information about various types of item scaling and IRT analysis has been presented by Van Schuur (2003). He began by elaborating a 'Guttman scale' and then introduced 'Rasch' and 'Mokken' Scale Analysis. Both Rasch and Mokken analysis assume random or probabilistic deviations from a perfect Guttman scale. Van Schuur stated that, in probabilistic IRT models, the probability of a positive response to a dichotomous item depends on one or more respondent parameters and one or more item parameters. The respondent and item parameters can be distinguished and estimated separately (cf. also Mokken, 1997; Molenaar & Sijtsma, 2000). If the probabilistic IRT model holds, the item parameters can be estimated separately from the respondents' scale values. This property is called 'specific objectivity' or 'respondent and item measurement invariance'. This scale characteristic makes it possible to compare respondents across groups or over time; it facilitates continuity between tests meant to measure the same concept at different but overlapping levels; and it supports computerised adaptive testing. A difference between Rasch and Mokken analysis is that Rasch analysis uses parametric logistic modelling, whereas Mokken analysis is based on nonparametric or ordinal modelling (Molenaar & Sijtsma, 2000). Rasch analysis allows item scoring on a continuous latent scale, whereas Mokken analysis uses addition of item scores to obtain a Mokken Scale sum score. Therefore, compared with Rasch analysis, Mokken analysis has less strict assumptions and can be used with a lower number of items (Van Schuur, 2003).

Given these potential advantages it is worthwhile to investigate violence assessment using a probabilistic IRT model, in particular the Mokken Scale model. Inspired by Marshall

and Hucker (2006), Nitschke, Osterheider, and Mokros (2009) have developed a Mokken scale based on 11 items to discriminate sexual sadists from sexual non-sadists. Nitschke et al. (2009) emphasised that research using Mokken unidimensional scaling is more efficient than research applying more dimensional scaling classifications. This issue can be explored empirically by performing secondary analysis of violence data on secondary school pupils collected by means of a national monitoring system (Mooij, 2011a, 2011b). In this research, teachers in the witness role, and pupils in the victim, offender, and witness roles, agreed perfectly in their frequency ranking of six types of violence (verbal, material, social, mild physical, severe physical, sexual). These results suggest that the different types of violence can be ordered more efficiently in terms of severity while indicating only one underlying construct. If this were shown to be true also for items indicating extreme violence (e.g., rape, using a weapon, stealing), then the resulting Mokken scale may well facilitate follow-up secondary analyses, assist in the design of future research and data analyses, and advance practical methods to be used by school personnel to actually reduce or prevent violence in and around schools. The research question is therefore as follows: Does Mokken Scale Analysis of secondary school pupils' experiences of violence result in a homogeneous, reliable, and valid unidimensional scale that fits all the requirements of Mokken scaling?

Method

Data and Samples

The research question will be answered by means of secondary analyses of Dutch national data collected in a digital survey of school safety ('school safety monitor'). The data are part of a two-yearly evaluation of school safety in secondary education (cf. Mooij, De Wit, & Fettelaar, 2011). The Internet-based survey was developed in 2006 to help the Ministry of

Education, Culture and Science record and evaluate trends in school safety among pupils, school staff, and school leadership. The results supply both national government and participating schools with cross-sectional and longitudinal information about school safety, experiences and incidents of violence, and school measures to reduce and prevent violence, at both national level and school level (cf. Mooij et al., 2011).

The secondary analysis makes use of some of the data collected in 2008 from a total of 78,840 pupils who participated in the survey. These pupils attended 219 secondary schools distributed throughout the Netherlands. They used an individual login code to complete a web questionnaire. The login code enabled randomised collection of different versions and parts of the questionnaire. The pupils completed the questionnaire in their classrooms under teacher supervision (see further Mooij, 2011a).

Two separate, random samples of 20% ($n=15,768$) were drawn from the total group of 78,840 pupil respondents. These samples were considered to represent the population of pupils adequately and to support efficiency in the Mokken analyses to be carried out. Moreover, the second sample offers an initial indication of the scale's external validity. The variables gender, country of birth, and feeling at home in the Netherlands were included as grouping variables to check the Mokken results for specific objectivity or for invariance of respondent and item measurement respectively. For reasons of analysis, pupils with one or more missing values were excluded. The first sample ($n_1=14,388$) was used to develop the desired scale and the second sample ($n_2=14,350$) functions as an independent check or cross-validation of the first result of scale construction.

Variables

Violence experienced at school is assessed by scoring different antisocial or aggressive activities related to verbal, material, social, mild physical, severe physical, and sexual

behaviours (Mooij, 2011a, 2011b). The specific items assessed with respect to each type of violence are given in Table 1.

Table 1 about here

All 29 items were scored for the September 2007 – January 2008 period. Scoring occurred by choosing one out of seven answer alternatives (from ‘never’ to ‘always’). The scores obtained for each violence item were dichotomised (never=0, once or more=1). Item scores per type of violent behaviour were included in principal factor analysis and Alpha scale analysis. For each type of violence, the factor results indicate the occurrence of a homogeneous group of items. Reliable Alpha scale coefficients on the dichotomised items of the six types of violence are presented in Table 2.

Table 2 about here

A robust Mokken scale must be relatively immune to variables potentially related to the subject of the scale. The first relevant variable here is gender: Boys or men experience other and more severe types of violence than girls or women (Marshall, 1992; Regan et al., 2006; Wang et al., 2010). Furthermore, the variables country of birth and feeling at home in the country in which one is actually living are shown to be important for the experience of violence in and around school (Beauvais & Jenson, 2002; Felix et al., 2009). Not being born in the country, and not feeling at home in the country, are related to experiences of more violent behaviour compared with being born and feeling at home in the country (Beauvais & Jenson, 2002). These three variables are included in the analyses in order to check invariance of measurement between groups when developing a Mokken scale. Gender is coded as ‘boy’ or ‘girl’, whereas country of birth and feeling at home are coded as ‘in the Netherlands=0’; ‘in another country=1’.

Mokken Scale Analysis (MSA)

Molenaar and Sijtsma (2000) and Van Schuur (2003) have introduced and elaborated on the nonparametric or ordinal Mokken Scale Analysis (MSA; cf. also Mokken, 1971; Nitschke et al., 2009; Sijtsma & Molenaar, 2002). These authors define and use different homogeneity statistics to specify the requirements for this type of scaling. First, the coefficient of homogeneity H_{ij} of a pair of items is expressed as the ratio of the covariance between items X_i and X_j and their maximum obtainable value, given the marginal distribution of these two items. H_{ij} thus indicates the internal consistency or reliability of each pair of items. Second, and likewise, the homogeneity of an entire scale, H , is the ratio of the sum of all pairwise covariances versus the sum of all pairwise maximum covariances, or the ratio of the sum of all errors observed versus expected. The H index then indicates the internal consistency of the whole scale. Third, H_i is the ratio of the sum of all pairwise covariances with respect to item i versus the sum of all pairwise maximum covariances concerning this item. H_i thus reflects the scalability of one item in relation to the whole set of items.

Statistics H , H_i , and H_{ij} are essential in constructing and testing the first aspect of a Mokken Scale. To control for chance values, the null hypothesis being tested is that H , H_i , or H_{ij} , are 0 in the population. A Mokken scale is said to exist if each $H_{ij} > 0$, and each $H_i > 0.30$. This furthermore implies that $H > 0.30$. If these three conditions are met, *monotone homogeneity* is said to occur. The interpretation is that a) the items reflect a unidimensional latent construct; b) local independence or respondent and item measurement invariance is at stake; and c) for each item it is true that, the more the respondent can be described in terms of the latent construct, the greater the chance that the response to the item will be positive ('person ordering is item-free').

If a scale is accepted as being monotone homogeneous, the set of items needs to be checked further for *double monotonicity*. Double monotonicity requires that the ordering of

items should be uniform across groups of respondents or, in other words, that item response functions do not intersect ('item ordering is person-free'). Molenaar and Sijtsma (2000) define and use coefficient H^T to test whether curves do intersect. H^T indicates the degree to which item ordering is the same for each respondent in the population. Negative values of H^T point to violations of the non-intersection requirement. The criterion is that the number of negative H^T values should not be larger than 10% of the total number of respondents; simultaneously, H^T for the whole group should be greater than 0.30. To find out which items intersect, Molenaar and Sijtsma (2000) use the $P^{I,I}$ matrix (see also Sijtsma & Molenaar, 2002). This is a square symmetrical matrix with items ordered according to their scores on the latent construct. The cell elements contain the proportion of respondents who give a positive response to both items in the pair. These positive responses are hypothesised to increase from left to right and from top to bottom, which indicates double monotonicity.

During the scale construction process, items violating scale requirements should be discarded. Van Schuur (2003) proposes removing items one at a time, to check the effects on the other items. While referring to Molenaar & Sijtsma (2000), he uses H_i in a combination of different indicators to develop statistic *Crit*. Each time, the item with the highest *Crit* value should be discarded first; *Crit* values ≤ 80 can be accepted. Finally, Van Schuur (2003) bases reliability ρ of a whole set of items in a Mokken Scale, or *Reliability Rho*, on the diagonal cells of the $P^{I,I}$ matrix (cf. also Molenaar & Sijtsma, 1988; Sijtsma & Molenaar, 1987). By interpolation, these cells estimate the probability of a positive response being given to the same item if it were completed twice.

MSA uses a bottom-up hierarchical clustering procedure, starting with the pair of items with the highest H_{ij} . The next best item in the scale is then included, and so on. MSA can be carried out with respect to dichotomous and polytomous items (Van Schuur, 2003) by using the computer program for 'Mokken Scaling Analysis for Polytomous items' (MSP) (see

Molenaar & Sijtsma, 2000). This program was used in the present analyses. However, the calculation of double monotonicity for polytomous items, or H^T , has not been implemented in this program. MSA was therefore conducted on items that were made dichotomous.

Results

Scale Construction using MSP (Sample 1)

Applying the MSP program with respect to all 29 violence items listed in Table 1 demonstrates monotone homogeneity. However, the *Crit* value of item 10 ('ignoring') is 150, which is above 80 and therefore evidently blocks double monotonicity. Elimination of this item followed by MSA of the remaining 28 items resulted in item 19 being discarded ('punching someone on purpose'). In successive analyses, items 11 ('excluding') and item 21 ('fighting with someone') were also eliminated. The MSA results for the remaining 25 items are shown in Table 3.

Table 3 about here

Table 3 shows that the lowest H_i value is 0.50 (item 25 'making sexual comments'), which is higher than the required minimum value 0.30. The H value of the whole scale is 0.56, exceeding the required minimum value 0.30 and indicating a strong scale (Molenaar & Sijtsma, 2000, p. 12). The H_{ij} values (not shown in Table 3) vary from 0.34 to 0.88, which are all above the required value 0. Taken together, the values of the three homogeneity coefficients demonstrate monotone homogeneity.

In addition, the highest *Crit* value is 33 (item 8 'destroying things'), which is below 40. Coefficient H^T is 0.54, which is above 0.30, whereas the number of negative H^T values amounts to 2.7%, i.e. below 10%. These scale results therefore meet the requirements for double monotonicity. The overall reliability of the scale is 0.94.

The scale results in Table 3 thus fulfil the Mokken scale requirements of both monotone homogeneity and double monotonicity. The 25 items are arranged from a low degree of violence ('calling someone names') to a high degree of violence ('rape'): see the items and their respective means in Table 3. This ordinal Mokken scale then assesses a latent construct reflecting pupils' experience of violent behaviour in terms of severity.

The next step of the analysis involves checking the scale results of Table 3 to clarify whether scaling is invariant when measuring different groups of respondents. The dichotomous variables gender, country of birth, and feeling at home in the Netherlands are used to differentiate between the pupils. The distributions of these variables within random sample 1 are presented in Table 4. Gender is distributed about equally within the sample, whereas having the Netherlands as the country of birth and feeling at home in the Netherlands both score above 90%.

Table 4 about here

Conducting MSA with respect to all six groups of pupils identified in sample 1 resulted in small differences in statistical indicators compared with the overall scale. However, within each subgroup the scale results are in line with MSA requirements. Summary statistics of the overall Mokken scale and the six subgroup Mokken scales are given in Table 5.

Table 5 about here

Validating the Mokken Scale with Sample 2 Data

An independent check of the exploratory results of random sample 1 can be carried out by applying MSA with respect to the data taken from random sample 2. The results of the cross validation show that the same four items discarded in sample 1 are also discarded in sample 2 in successive steps. An overview of the MSA results is given in Table 6. The table shows that H is 0.55; H_i values are at least 0.49, which exceeds 0.30; and H_{ij} values range from 0.33 to

0.89. Sample 2 outcomes thus meet the requirements of monotone homogeneity. Generally speaking, the MSA sample 2 results in Table 6 closely match the MSA sample 1 results in Table 3.

Table 6 about here

Items 13 and 14 produce the two highest *Crit* values, 23 and 25, which do not exceed the level of acceptability (80). Coefficient H^T for the whole sample is 0.51 and the percentage of negative H^T values is 2.5. The results for sample 2 in Table 6 therefore also meet the requirements of double monotonicity. A minor difference between sample 1 results and sample 2 results is the ordering of item 8 ('destroying things') and item 14 ('blackmailing'). In Table 3, the order of these two items is 8, 14; in Table 6 this is reversed. The difference between the respective means is negligible, however.

Like sample 1, Mokken scaling for sample 2 has to be invariant when measuring different subgroups. Table 4 provides information on the distribution characteristics of sample 2 regarding the dichotomous variables gender, country of birth, and feeling at home in the Netherlands. As in sample 1, the outcomes of MSA with respect to all six groups of sample 2 result in some minor differences compared with the overall Mokken scale. Within each subgroup, scale results are in line with MSA requirements. Table 7 summarises the statistics of the overall scale and the six subgroup scales for sample 2.

Table 7 about here

Discussion

Mokken Scale Analysis of secondary school pupils' information about their experiences of violence results in a homogeneous and reliable Mokken scale that meets the requirements of both monotone homogeneity and double monotonicity. This affirmative answer to the

research question demonstrates the existence of a unidimensional, cumulative scale for pupils' experiences of violence. Moreover, the first exploratory results of sample 1 (n=14,388) are cross-validated by the independent check carried out by sample 2 (n=14,820). The latent construct underlying the ordinal Mokken scale measuring pupils' experience of violence in terms of severity is composed of 25 items. These are ordered from 'occurring most frequently' (item 25, 'calling someone names') to 'occurring least frequently' (item 1, 'rape'): See Tables 3 and 6. This result implies that the six Alpha scales representing different types of violence (see Tables 1 and 2) can be replaced, if one so desires, by one Mokken scale indicating *experience of violence in terms of severity*. This latent construct varies from verbal and mild physical violence, which occurs relatively most frequently, to combinations of social, material and severe physical violence, to very severe and serious sexual violence, which occurs relatively least frequently.

This empirical ordering appears to be plausible and in line with what can be expected from secondary school pupils. Moreover, in both sample 1 and sample 2, this ordering demonstrates invariance between boys and girls; between those born in the Netherlands or elsewhere; and between those who feel at home in the Netherlands and those who do not. These verifications of measurement invariance for different groups support the applicability of the Mokken scale for different subgroups of respondents who are known to experience violence in varying degrees of severity. To further explore the present results, MSA was conducted using the 15 items that were chosen most frequently by a total of 74,260 pupils. Most of these items indicate verbal and mild physical types of violence, or lower levels of experience of violence in terms of severity (cf. Table 3). While using the same statistical criteria as before, the outcomes show that six items had to be removed in order to get an adequate Mokken Scale consisting of nine items. This scale combines verbally and physically

disturbing behaviour and clarifies that different Mokken scales can be obtained by analysing different sets of items.

Furthermore, the measurement invariance for different subgroups of respondents concerns the requirements of both monotone homogeneity and double monotonicity for each subgroup e.g. boys and girls (cf. Tables 5 and 7). MSA focuses on the consistency and unidimensionality of the cumulative order of items in a scale within boys and within girls. This gender invariance differs from evaluation by a structural equation model (SEM) analysis or multiple regression analysis in which gender acts as an ‘exogenous variable’ influencing a Mokken sum score. SEM and multiple regression analysis concentrate on analysis of potential differences with respect to between-gender variation in scoring on a latent construct or an overall Mokken scale score. For example, multiple regression analyses of these same data reveal among other things that boys score significantly higher than girls on the Mokken sum score concerning experience of violence in terms of severity (cf. Mooij, submitted).

One limitation of the study, however, is that the analysis had to be performed on items that were dichotomised. If it were available in MSP or another program, MSA involving polytomous items might throw more light on the consequences of working with dichotomised answer categories, as was necessary in the present analyses. Despite this limitation, the present Mokken results certainly support the approach used by Michie and Cooke (2006) and Nitschke et al. (2009) to overcome existing problems in violence measurement by applying Mokken Scale Analysis. The actual study results confirm the adequacy of their intentions and offer suggestions for future explorations of the relevancy and use of Mokken Scale Analysis in this type of research.

Given the national data used, the Mokken results can be generalised to the population of Dutch secondary school pupils (cf. Mooij et al., 2008). Generalisability to other groups of respondents, education systems, or countries should be based on adequate research, to

provide more information and empirical evidence. In this respect, one important advantage of the scaling results in Tables 3 – 7 is that the information on the 25 items can be transformed into one sum score estimating each respondent's experience of violence in terms of severity. This sum score can be used in follow-up secondary analyses, for example, in order to learn more about its meaning and practical uses (cf. Mooij, submitted). Furthermore, the score indicating the severity of violence experienced by a respondent can be included in a theoretical multilevel framework, to both explain and predict the severity of violence that a pupil may experience in an educational context. At individual level, the relevant variables are gender, educational attainment, social and emotional characteristics, and age; at class level, the relevant variables indicate teachers' social or disciplinary policy and curricular differentiation; and at school level, variables such as social norms, school cohesion, school responsiveness to violence, and severity of disciplinary policy interact in various ways with pupil-level and class-level variables, producing a fairly complete, evidence-based picture of interrelated processes and effects (cf. Kettler, 2011; Mooij, Smeets, & De Wit, 2011; You, Ritchey, Furlong, Shochet, & Boman, 2011).

Acknowledgements

The national school safety monitor survey for secondary education was developed at the request of the Dutch Ministry of Education, Culture and Science, which consented to secondary analyses of the data and publication of the results. I am grateful to Rick Schaap, Halewijn Drent and Daan Fettelaar MSc for their assistance in performing the necessary analyses.

References

- Bayh, B. (1975). *Our nation's schools – a report card: "A" in school violence and vandalism*. Washington, DC: US Government Printing Office.
- Beauvais, C., & Jenson, J. (2002). *Social Cohesion: Updating the state of the research. CPRN Discussion Paper*. Ottawa, Canada: Canadian Policy Research Networks. Retrieved from http://www.cprn.com/documents/12949_en.pdf
- Chen, J., & Astor, R.A. (2011). School engagement, risky peers, and student-teacher relationships as mediators of school violence in Taiwanese vocational versus academically oriented high schools. *Journal of Community Psychology*, 39, 10-30.
- Duke, N.N., Pettingell, S.L., McMorris, B.J., & Borowsky, I.W. (2010). Adolescent violence perpetration: Associations with multiple types of adverse childhood experiences. *Pediatrics*, 125, 778-786.
- Felix, E.D., Furlong, M.J., & Austin, G. (2009). A cluster analytic investigation of school violence victimization among diverse students. *Journal of Interpersonal Violence*, 24(10), 1673-1695.
- Finkelhor, D., Omrod, R., Turner, H., & Hamby, S. (2011). School, police, and medical authority involvement with children who have experienced victimization. *Archives of Pediatrics and Adolescent Medicine*, 165, 9-15.
- Kettler, R.J. (2011). Computer-based screening for the new modified alternate assessment. *Journal of Psychoeducational Assessment*, 29(1), 3-13. doi:10.1177/0734282910370804
- Marshall, L.L. (1992). The Severity of Violence Against Men Scales. *Journal of Family Violence*, 7(3), 189-203.
- Marshall, W.L., & Hucker, S.J. (2006). Severe sexual sadism: Its features and treatment. In R.D. McAnulty and M.M. Burnette (Eds.), *Sex and sexuality. Sexual deviation and sexual offenses* (Vol. 3, pp. 227-250). Westport, CT: Praeger.

- Mayer, M.J., & Furlong, M.J. (2010). How safe are our schools? *Educational Researcher*, 39(1), 16-26.
- Mayer, M.J. & Leone, P.E. (1999). A structural analysis of school violence and disruption: Implications for creating safer schools. *Education and Treatment of Children*, 22, 333-356.
- Michie, C., & Cooke, D.J. (2006). The structure of violent behaviour: A hierarchical model. *Criminal Justice and Behavior*, 33(6), 706-737.
- Ministry of Education, Culture and Science. (2011). *Kamerbrief cijfers drugsgebruik op scholen [Informing Parliament about the use of drugs at school]*. The Hague, The Netherlands: Author.
- Moffitt, T.E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*, 100, 674-701.
- Mokken, R.J. (1971). *A theory and procedure of scale analysis*. The Hague, The Netherlands: Mouton.
- Mokken, R.J. (1997). Nonparametric models for dichotomous responses. In W.J. van der Linden and R.K. Hambleton (Eds.), *Handbook of Modern Item Response Theory* (pp. 351-367). New York: Springer.
- Molenaar, I.W., & Sijtsma, K. (1988). Mokken's approach to reliability estimation extended to multcategory items. *Kwantitatieve Methoden*, 9(28), 115-126.
- Molenaar, I.W., & Sijtsma, K. (2000). *MSP5 for Windows: A Program for Mokken Scale Analysis For Polytomous Items (Version 5.0)*. Groningen, The Netherlands: University of Groningen, ProGAMMA.
- Mooij, T. (2011a). Differences in pupil characteristics and motives in being a victim, perpetrator and witness of violence in secondary education. *Research Papers in Education*, 26, 105-128.

- Mooij, T. (2011b). Secondary school teachers' personal and school characteristics, experience of violence and perceived violence motives. *Teachers and Teaching: theory and practice*, 17(2), 227-253.
- Mooij, T. (submitted). Explaining severity of violence experienced by secondary school pupils.
- Mooij, T., De Wit, W., & Fettelaar, D. (2011). *Sociale veiligheid in en rond scholen. Primair en secundair onderwijs 2006 – 2010 [Social safety in and around schools. Primary and secondary education 2006-2010]*. Nijmegen, The Netherlands: Radboud Universiteit, ITS.
- Mooij, T., Smeets, E., & De Wit, W. (2011). Multi-level aspects of social cohesion of secondary schools and pupils' feelings of safety. *British Journal of Educational Psychology*, 81(3), 369-390. doi:10.1348/000709910X526614
- Mynard, H.M., & Joseph, S. (2000). Development of the multidimensional peer-victimization scale. *Aggressive Behavior*, 26, 169-178.
- Nitschke, J., Osterheider, M., & Mokros, A. (2009). A cumulative scale of severe sexual sadism. *Sexual Abuse: A Journal of Research and Treatment*, 21(3), 262-278.
- Nylund, K., Bellmore, A., Nishina, A., & Graham, S. (2007). Subtypes, severity, and structural stability of peer victimization: What does latent class analysis say? *Child Development*, 78(6), 1706-1722.
- Office of the Children's Commissioner. (2010). *Report of Children's Commissioner for the year ended 30 June 2010*. Ministry for Social Development and Employment, New Zealand: Author. Retrieved from http://www.occ.org.nz/__data/assets/pdf_file/0010/7858/CC_Annual_Report_2010.pdf
- Regan, K.V., Bartholomew, K., Kwong, M.J., Trinke, S.J., & Henderson, A.J.Z. (2006). The relative severity of acts of physical violence in heterosexual relationships: An item response theory analysis. *Personal Relationships*, 13, 37-52.

- Schafer, J. (1996). Measuring spousal violence with the Conflict Tactics Scale. *Journal of Interpersonal Violence, 11*, 572-585.
- Schwartz, D., Gorman, A.H., Nakamoto, J., & Toblin, R.L. (2005). Victimization in the peer group and children's academic functioning. *Journal of Educational Psychology, 97*(3), 425-435.
- Sijtsma, K., & Molenaar, I.W. (1987). Reliability of test scores in nonparametric item response theory. *Psychometrika, 52*, 79-97.
- Sijtsma, K., & Molenaar, I.W. (2002). *Introduction to nonparametric item response theory*. Thousand Oaks, CA: Sage.
- Siu, A. F. Y. (2011). Validation of the Substance Use Risk Profile Scale for adolescents in Hong Kong. *Journal of Psychoeducational Assessment, 29*(1), 75-83.
doi:10.1177/0734282910362044
- Solberg, M.E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior, 29*, 239-268.
- Van Schuur, W.H. (2003). Mokken Scale Analysis: Between the Guttman Scale and Parametric Item Response Theory. *Political Analysis, 11*, 139-163.
- Wang, J., Iannotti, R.J., Luk, J.W., & Nansel, T.R. (2010). Co-occurrence of victimization from five subtypes of bullying: Physical, verbal, social exclusion, spreading rumors, and cyber. *Journal of Pediatric Psychology, 35*, 1103-1112.
- You, S., Ritchey, K.M., Furlong, M.J., Shochet, I., & Boman, P. (2011). Examination of the latent structure of the Psychological Sense of School Membership Scale. *Journal of Psychoeducational Assessment, 29*(3), 225-237. doi:10.1177/0734282910379968

Table 1

Types of Violent Behaviour and Specification into Items

<i>Types of violence</i>	<i>Specification into items</i>
Verbal	Calling someone names, bothering someone on purpose, talking in an extra loud voice, making a lot of noise on purpose
Material	Scratching or damaging something, spray-painting or dirtying something, hiding or mislaying something, destroying things, stealing
Social	Ignoring, excluding, threatening, intimidating, blackmailing, spreading false rumours
Mild physical	Striking or hurting someone on purpose, pushing or kicking someone on purpose, tripping someone on purpose, punching someone on purpose, hitting
Severe physical	Fighting with someone, beating or roughing someone up, threatening someone with a weapon, using a weapon
Sexual	Making sexual comments, sexual gestures, directed sexual contact or touch, sexually molesting someone, rape

Table 2

Alpha Reliabilities of Scales on Types of Violent Pupil Behaviour

<i>Scales on violent behaviour</i>	<i>n items</i>	<i>Cronbach's α</i>
Verbal	4	.85
Material	5	.83
Social	6	.81
Mild physical	5	.89
Severe physical	4	.72
Sexual	5	.78

Table 3

*Sample 1: Mokken Violence Scale of 25 Items (Items 10, 19, 11 and 21 eliminated)**

<i>Cumulative order</i>	<i>Type of violence</i>	<i>Item</i>	<i>Label</i>	<i>Mean</i>	<i>H_i</i>	<i>Restscore (Crit)**</i>
1	Sexual	29	rape	0.04	0.66	
2	Severe phys.	24	using a weapon	0.05	0.68	
3	Sexual	28	sexually molesting someone	0.06	0.61	
4	Severe phys.	23	threatening someone with a weapon	0.07	0.66	
5	Sexual	27	directed sexual contact or touch	0.13	0.57	
6	Material	6	spray-painting or dirtying something	0.17	0.55	
7	Material	9	stealing	0.18	0.55	
8	Severe phys.	22	beating or roughing someone up	0.21	0.52	
9	Social	12	threatening	0.23	0.54	
10	Social	13	intimidating	0.24	0.55	
11	Material	8	destroying things	0.25	0.56	33
12	Social	14	blackmailing	0.25	0.52	28
13	Material	5	scratching or damaging something	0.31	0.51	
14	Sexual	26	sexual gestures	0.32	0.52	
15	Material	7	hiding or mislaying something	0.36	0.51	
16	Sexual	25	making sexual comments	0.40	0.50	
17	Social	15	spreading false rumours	0.41	0.51	
18	Mild phys.	20	hitting	0.44	0.54	
19	Mild phys.	18	tripping someone on purpose	0.45	0.56	
20	Mild phys.	17	pushing or kicking someone on purpose	0.50	0.60	
21	Mild phys.	16	striking or hurting someone on purpose	0.53	0.60	
22	Verbal	4	making a lot of noise on purpose	0.64	0.59	
23	Verbal	2	bothering someone on purpose	0.65	0.60	
24	Verbal	3	talking in an extra loud voice	0.66	0.61	
25	Verbal	1	calling someone names	0.78	0.65	
N=14,388; Coefficient $H=0.56$; $\rho=0.94$						
Matrix of H_{ij} values per item pair: minimum 0.34; maximum 0.88						
H^T coefficient for entire group: 0.54						
Number of negative H^T values: 341 (2.7%)						

* Primary data were polytomous; data were dichotomised in Mokken Scale Analysis.

** Only the two highest values are given.

Table 4

Sample Distributions according to Gender, Country of Birth, and Feeling at Home

		<i>Boys</i>		<i>Other country of birth</i>		<i>Feel at home in other country</i>	
<i>Total N</i>		<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Sample 1	14,388	7,100	49.3	785	5.5	1339	9.3
Sample 2	14,350	7,348	51.2	739	5.1	1364	9.5

Table 5

Sample 1: Summary Statistics for Gender, Country of Birth, and Feeling at Home in Country

	<i>Overall</i>	<i>Gender</i>		<i>Country of birth</i>		<i>Feeling at home</i>	
	<i>scale</i>	<i>Boys</i>	<i>Girls</i>	<i>Nthlds</i>	<i>Other</i>	<i>Nthlds</i>	<i>Other</i>
<i>Total group N = 14,388</i>							
<i>Monotone homogeneity</i>							
Coefficient H	0.56	0.57	0.54	0.56	0.58	0.55	0.61
ρ	0.94	0.94	0.93	0.94	0.95	0.94	0.96
Matrix of H_{ij} values per item pair							
minimum	0.34	0.33	0.31	0.34	0.33	0.34	0.42
maximum	0.88	0.91	0.88	0.89	0.97	0.89	0.90
<i>Double monotonicity</i>							
H^T coefficient for entire group	0.54	0.51	0.52	0.52	0.44	0.52	0.46
Negative H^T values							
number	341	199	151	325	21	301	47
percentage	2.7	3.1	2.3	2.7	3.1	2.6	4.1

Table 6

Sample 2: Cross Validation of Mokken Scale (cf. Table 3)

<i>Cumulative order</i>	<i>Item</i>	<i>Label</i>	<i>Mean</i>	<i>H_i</i>	<i>Restscore (Crit)*</i>
1	29	rape	0.04	0.65	
2	24	using a weapon	0.05	0.66	
3	28	sexually molesting someone	0.06	0.61	
4	23	threatening someone with a weapon	0.06	0.64	
5	27	directed sexual contact or touch	0.13	0.55	
6	6	spray-painting or dirtying something	0.17	0.55	
7	9	stealing	0.19	0.52	
8	22	beating or roughing someone up	0.21	0.50	
9	12	threatening	0.22	0.54	
10	13	intimidating	0.24	0.55	23
11	14	blackmailing	0.24	0.52	25
12	8	destroying things	0.25	0.55	
13	5	scratching or damaging something	0.31	0.51	
14	26	sexual gestures	0.32	0.51	
15	7	hiding or mislaying something	0.36	0.51	
16	25	making sexual comments	0.40	0.49	
17	15	spreading false rumours	0.41	0.50	
18	20	hitting	0.44	0.52	
19	18	tripping someone on purpose	0.45	0.55	
20	17	pushing or kicking someone on purpose	0.50	0.59	
21	16	striking or hurting someone on purpose	0.53	0.59	
22	4	making a lot of noise on purpose	0.64	0.59	
23	2	bothering someone on purpose	0.65	0.60	
24	3	talking in an extra loud voice	0.66	0.61	
25	1	calling someone names	0.79	0.65	

N=14,820; Coefficient $H=0.55$; $\rho=0.94$ Matrix of H_{ij} values per item pair: minimum 0.33; maximum 0.89 H^T coefficient for entire group: 0.51Number of negative H^T values: 333 (2.5%)

* Only the highest values are given.

Table 7

Sample 2: Summary Statistics for Gender, Country of Birth, and Feeling at Home in Country

	Overall	Gender		Country of birth		Feeling at home	
	scale	Boys	Girls	Nthlds	Other	Nthlds	Other
<i>Total group N = 14,820</i>							
<i>Monotone homogeneity</i>							
Coefficient H	0.55	0.55	0.54	0.55	0.58	0.54	0.59
ρ	0.94	0.94	0.93	0.94	0.95	0.94	0.95
Matrix of H_{ij} values per item pair							
minimum	0.33	0.33	0.31	0.33	0.34	0.32	0.37
maximum	0.89	0.91	0.88	0.89	0.95	0.89	0.95
<i>Double monotonicity</i>							
H^T coefficient for entire group	0.51	0.51	0.52	0.52	0.42	0.52	0.44
Negative H^T values							
number	333	192	135	309	30	275	57
percentage	2.5	2.8	2.1	2.4	4.4	2.3	4.9