

Ten-year trends in adolescents' self-reported emotional and behavioral problems in the Netherlands

Elisa L. Duinhof · Gonneke W. J. M. Stevens ·
Saskia van Dorsselaer · Karin Monshouwer ·
Wilma A. M. Vollebergh

Received: 15 July 2014 / Accepted: 11 December 2014 / Published online: 23 December 2014
© Springer-Verlag Berlin Heidelberg 2014

Abstract Changes in social, cultural, economic, and governmental systems over time may affect adolescents' development. The present study examined 10-year trends in self-reported emotional and behavioral problems among 11- to 16-year-old adolescents in the Netherlands. In addition, gender (girls versus boys), ethnic (Dutch versus non western) and educational (vocational versus academic) differences in these trends were examined. By means of the Strengths and Difficulties Questionnaire, trends in emotional and behavioral problems were studied in adolescents belonging to one of five independent population representative samples (2003: $n = 6,904$; 2005: $n = 5,183$; 2007: $n = 6,228$; 2009: $n = 5,559$; 2013: $n = 5,478$). Structural equation models indicated rather stable levels of emotional and behavioral problems over time. Whereas some small changes were found between different time points, these changes did not represent consistent changes in problem levels. Similarly, gender, ethnic and educational differences in self-reported problems on each time point were highly comparable, indicating stable mental health inequalities between groups of adolescents over time. Future internationally comparative studies using multiple measurement moments are needed to monitor whether these persistent mental health inequalities hold over extended periods of time and in different countries.

Keywords Time-trends · Adolescents · Emotional and behavioral problems · Gender, ethnic and educational differences

Introduction

Recent decades have shown ample societal changes in Western countries with a potential positive or negative impact on the lives of adolescents (e.g., [11, 12, 24, 28]). By means of time-trend studies on adolescent emotional and behavioral problems, possible effects of changing societies on the lives of adolescents can be detected. To examine trends in adolescents' emotional and behavioral problems in a reliable manner, repeated assessments with comparable measures in population representative samples are needed [24]. Studies meeting these criteria are relatively rare and show mixed findings. In both Europe and the USA, most studies reported increased population levels of adolescents' emotional problems over time (e.g., [3, 13, 14, 19, 33, 34, 37, 43]). In addition, some studies found increased population levels of behavioral problems [3, 13]. In contrast, other studies reported slightly decreased population levels of behavioral problems [35] or both (slightly) decreased emotional and behavioral population levels [2, 23] over time. Furthermore, some time-trend studies found stable population levels of emotional and behavioral problems [31, 40, 41]. Although the former might roughly suggest that levels of emotional problems have increased during the last decades whereas this may not be the case for behavioral problems the findings of previous research are somewhat inconsistent. As such, it is difficult to draw firm conclusions about trends in adolescents' emotional and behavioral problems in Western countries.

E. L. Duinhof · G. W. J. M. Stevens (✉) · W. A. M. Vollebergh
Utrecht Centre for Child and Adolescent Studies, Utrecht
University, P.O. Box 80.140, 3508 TC Utrecht, The Netherlands
e-mail: g.w.j.m.stevens@uu.nl

S. van Dorsselaer · K. Monshouwer
Trimbos Institute (Netherlands Institute of Mental Health
and Addiction), P.O. Box 725, 3500 AS Utrecht, The Netherlands

In addition, previous studies on time trends in emotional and behavioral problems were limited because they only used two or three measurement moments with approximately 10-year time intervals. Therefore, the results of these studies might not reflect consistent trends in emotional and behavioral problems over time, but more or less random problem levels at different points of time. Finally, societal changes may have a differential impact on emotional and behavioral problems of different social groups. The impact of societal changes on the daily lives of adolescents may not only vary with the social group, but these daily life changes may also have differential effects on emotional and behavioral problems for different social groups [28]. This suggests that different time-trends in emotional and behavioral problems may be found for different social groups. In this study, it is tested whether time-trends in emotional and behavioral problems vary with the gender of the adolescent as well as with their ethnic and educational background. Some studies were indicative of gender-specific trends in emotional and behavioral problems, with girls showing increasing population levels of emotional problems and boys showing stable or decreasing population levels of emotional problems over time [35, 43]. However, although previous studies on time trends stratified by or controlled for social group differences (i.e., gender, ethnic minority status), these did not empirically test whether these trends vary with the gender, ethnic minority status or educational level of the adolescent.

To find out if adolescents' mean levels of emotional and behavioral problems have been subjected to change during the last decade, the present study aimed to investigate trends in Dutch adolescents' self-reported emotional and behavioral problems between 2003 and 2013 as well as gender, ethnic, and educational differences in these trends. To do so, five nationally representative samples were used measuring emotional and behavioral problems on consecutive time points.

Materials and methods

Samples

Five independent population representative samples of the Dutch Health Behavior in School-aged Children (HBSC; 2005, 2009, 2013) [15–17] and the Dutch National School Survey on Substance Use (DNSSSU; 2003, 2007) [25, 26] studies were used. The total sample consisted of 29,352 11- to 16-year-old boys and girls attending secondary education. Consent procedures required by ethical and legal authorities for this type of survey were followed; only those adolescents who volunteered to participate and whose parents did not object to their participation were included in the current study.

The HBSC and the DNSSSU studies made use of a highly comparable data collection and sampling methods and were conducted by the same national research team. In both studies, field work took place in October and November and samples were drawn using cluster sampling, with schools as the primary sampling units. All samples were stratified by degree of urbanization. In addition, data were weighted for school year, educational level, degree of urbanization and gender to assure that all samples were nationally representative. All data were collected by means of questionnaires, which were distributed in school classes and administered by research assistants during a lesson (usually 50 min). When introducing the questionnaire, the anonymity of the answers of the participants was emphasized by the research assistant. Collecting all questionnaires in one envelope and sealing the envelope in the presence of the participants further ensured anonymity. Also, during administration of the questionnaires, research assistants ensured that participants were not able to read each other's answers. More information about the data collection procedure can be found elsewhere (e.g., [15, 17, 25]).

The average school response rate was 52.8 %, ranging from 40 % (2013) to 72 % (2003). Main reasons for schools not to participate in the study were participation in, or frequent research participation requests by other surveys. Within randomly selected school classes, in all samples, response rates were above 91 %. The main reasons for non-response of students were illness or truancy. Of the total sample, for 900 adolescents (2003: $n = 162$, 2.3 %; 2005: $n = 238$, 4.4 %; 2007: $n = 328$, 5.0 %; 2009: $n = 81$, 1.4 %; 2013: $n = 91$, 1.6 %) all items assessing emotional and behavioral problems were missing and the adolescents were excluded from the analyses. Comparing these adolescents with the remaining samples, some significant ($p < 0.05$) but negligible ($\varphi < 0.10$) gender, ethnic and educational differences were found. For the remaining samples, missing item responses varied between 0.1 and 2.3 %. Although for some items missing responses were significantly ($p < 0.05$) related to participants' characteristics (i.e., gender, ethnicity and educational level), these relationships were all negligible ($\varphi < 0.10$) in size.

Instruments

Emotional and behavioral problems

All participants filled in the Dutch translation of the self-report version of the Strengths and Difficulties Questionnaire (SDQ) [20]. The self-report SDQ is a 25-item behavioral screening questionnaire, for adolescents from 11- to 16-years that has been used throughout the Western world [1, 20]. The SDQ consists of five subscales, each including five items, measuring difficulties (conduct problems,

emotional problems, hyperactivity, peer problems) and strengths (prosocial behavior). Items are scored on a three-point ordinal Likert scale ('Not true' = 0, 'Somewhat true' = 1; 'Certainly true' = 2), with higher scores indicating more problems. Examples of items are: 'I worry a lot' and 'I am easily distracted, I find it difficult to concentrate'. The items that assessed difficulties were included in this study. For descriptive purposes subscale mean scores were computed. Mean scores were only computed for participants who completed at least three of the five subscale items.

In the present study, Cronbach's alphas were acceptable for emotional problems ($\alpha_{\text{range}} = 0.67\text{--}0.71$) and hyperactivity ($\alpha_{\text{range}} = 0.68\text{--}0.73$). However, in line with former studies (e.g., [44]), unacceptably low values were found for conduct problems ($\alpha_{\text{range}} = 0.48\text{--}0.54$) and peer problems ($\alpha_{\text{range}} = 0.42\text{--}0.49$). Moreover, the construct validity of the SDQ has been debated upon (e.g., [21]). Therefore, the construct validity of the SDQ was examined in the current sample before proceeding with subsequent analyses (see "Results").

Demographic variables

Age, gender, ethnicity (i.e., native Dutch vs. non western background) and educational level (i.e., vocational vs. academic) were all measured by means of self-report. The academic educational level consisted of adolescents following pre-university education, higher secondary education and intermediate secondary education. These educational levels were compared with the vocational track. Adolescents were considered as having a non western ethnicity if at least one of the parents of the adolescent was born in a non western country.

Analytic strategy

Structural Equation Modeling analyses were performed in Mplus 7.11, using the weighted least squares mean and variance adjusted estimator. Analyses were corrected for cluster effects of adolescents sharing the same school. First, using multigroup confirmatory factor analysis, the factor structure of the SDQ with four latent factors representing emotional and behavioral problems was examined. Second, it was tested whether the SDQ measured identical underlying constructs in each sample (i.e., time point), and subgroup (i.e., gender, ethnicity, and educational level). To do so, analyses of invariance were performed by testing three consecutive models: a model without equality constraints (i.e., configural invariance) a model with equally constrained factor loadings (i.e., metric invariance), and a model with equally constrained factor loadings and item thresholds (i.e., scalar invariance). Measurement invariance

was established if models showed acceptable fits to the data, no significant increases in Chi square ($\Delta\chi^2$) and negligible differences in model fit (ΔCFI and ΔRMSEA) compared to the configural model. CFI values above 0.900 and RMSEA values below 0.080 suggested acceptable model fits (for details and references see [7]). Values not exceeding 0.010 for ΔCFI and 0.015 for ΔRMSEA were considered negligible [8].

Second, latent mean differences in problems between time points were examined using a multigroup model. The Wald test was used to examine whether latent means were comparable across all five time points. In case the Wald test reached significance, indicating that values were not comparable across all five time points, latent mean differences between successive time points were examined. This was investigated by successively fixing latent means at zero for each time point because multigroup models require latent means to be fixed at zero for one group. To examine the size of latent mean differences between time points the latent standardized effect size for independent samples (D) [9] was used.

Third, gender, ethnic and educational differences in trends were examined. To test this, multigroup Multiple Indicators Multiple Causes (MIMIC) models were constructed by regressing the dichotomized variables gender, ethnicity and educational level on the latent factors. The Wald test was used to examine whether subgroup mean differences changed significantly across all five time points. To examine the size of group mean differences on each time point, standardized regression coefficients were interpreted using Cohen's d [negligible ($d < 0.20$), small ($d < 0.50$), medium ($d < 0.80$), large ($d > 0.80$)] [10].

Results

Descriptives

The key demographic statistics of the five samples are presented in Table 1. Little age, gender, ethnic and educational level differences were found between the samples. Although samples differed significantly in their age, ethnic and educational level composition, these differences were all negligible in size. In addition, Table 1 shows that, with subscale scores ranging from 0 to 10, mean levels of conduct problems, emotional problems and peer problems were rather low on all time points. For hyperactivity, moderate mean levels were found on all time points.

Tests of measurement invariance

Results of the multigroup confirmatory factor analysis indicated that the four factor multigroup model showed an

Table 1 Demographic and descriptive statistics of the samples (weighted %)

Survey year ^a	DNSSSU 2003	HBSC 2005	DNSSSU 2007	HBSC 2009	HBSC 2013
Total, <i>n</i>	6,904	5,183	6,228	5,559	5,478
Age, M(SD)	13.84 (1.27)	13.86 (1.27)	13.89 (1.25)	13.85 (1.27)	13.83 (1.28)
Gender, <i>n</i> (%)					
Girls	3,376 (48.9)	2,550 (49.2)	3,054 (49.0)	2,733 (49.2)	2,694 (49.2)
Boys	3,528 (51.1)	2,633 (50.8)	3,174 (51.0)	2,826 (50.8)	2,784 (50.8)
Ethnicity, <i>n</i> (%)					
Dutch	5,443 (84.1)	3,902 (80.3)	5,059 (86.9)	4,442 (83.8)	4,359 (84.5)
Non western	1,025 (15.9)	956 (19.7)	765 (13.1)	859 (16.2)	799 (15.5)
Educational level, <i>n</i> (%)					
Vocational	1,829 (26.9)	1,481 (28.7)	1,407 (22.6)	1,124 (20.2)	1,335 (24.4)
Academic	4,958 (73.1)	3,680 (71.3)	4,821 (77.4)	4,435 (79.8)	4,143 (75.6)
Self-reported problems, M(SD)					
Conduct problems	2.24 (1.63)	2.01 (1.55)	1.95 (1.59)	1.83 (1.50)	1.83 (1.51)
Emotional problems	2.55 (2.14)	2.17 (2.07)	2.46 (2.09)	2.16 (2.00)	2.53 (2.20)
Hyperactivity	4.15 (2.22)	3.57 (2.19)	3.78 (2.28)	3.77 (2.24)	3.93 (2.31)
Peer problems	1.80 (1.67)	1.71 (1.62)	1.71 (1.62)	1.47 (1.53)	1.57 (1.58)

^a No differences in gender composition ($\chi^2(4) = 0.16$, $p = 0.997$, $\varphi = 0.002$), and significant but negligible differences in ethnicity ($\chi^2(4) = 85.92$, $p < 0.001$, $\varphi = 0.056$), educational level ($\chi^2(4) = 138.19$, $p < 0.001$, $\varphi = 0.069$), and age ($F(4) = 11.48$, $p < 0.001$, $\eta^2 = 0.002$) were found between samples

Table 2 Fit indices of the models testing measurement invariance and multigroup MIMIC models

Models ^a	$\chi^2/\Delta\chi^2$	<i>df</i> / Δ <i>df</i>	CFI/ Δ CFI	RMSEA/ Δ RMSEA	90 % CI of RMSEA
Measurement invariance over time					
1. Configural invariance	9,136.22*	815	0.890	0.042	0.041–0.042
2. Metric invariance	7,939.06*	879	0.907	0.037	0.036–0.038
3. Scalar invariance	8,855.31*	943	0.896	0.038	0.037–0.039
Difference model 1 and 2	237.76*	64	0.017	0.005	
Difference model 1 and 3	726.61*	128	0.006	0.004	
Measurement invariance gender					
4. Configural invariance	10,310.38*	326	0.893	0.046	0.045–0.046
5. Metric invariance	9,298.74*	342	0.904	0.042	0.042–0.043
6. Scalar invariance	10,285.07*	358	0.894	0.043	0.043–0.044
Difference model 4 and 5	207.14*	16	0.011	0.004	
Difference model 4 and 6	737.68*	32	0.001	0.003	
Measurement invariance ethnicity					
7. Configural invariance	8,678.93*	326	0.895	0.043	0.042–0.044
8. Metric invariance	7,537.48*	342	0.909	0.039	0.038–0.040
9. Scalar invariance	7,936.06*	358	0.905	0.039	0.038–0.040
Difference model 7 and 8	97.07*	16	0.014	0.004	
Difference model 7 and 9	200.79*	32	0.010	0.004	
Measurement invariance educational level					
10. Configural invariance	9,439.62*	326	0.893	0.044	0.043–0.045
11. Metric invariance	8,405.75*	342	0.905	0.040	0.039–0.041
12. Scalar invariance	8,866.98*	358	0.900	0.040	0.040–0.041
Difference model 10 and 11	188.12*	16	0.012	0.004	
Difference model 10 and 12	271.29*	32	0.007	0.004	
Multigroup MIMIC models					
13. Gender	9,163.72*	1,023	0.895	0.037	0.036–0.038
14. Ethnicity and educational level	8,512.64*	1,103	0.891	0.035	0.034–0.036

* $p < 0.001$

^a To increase global model-fit, the first and second item measuring hyperactivity were allowed to correlate based on similar meaning

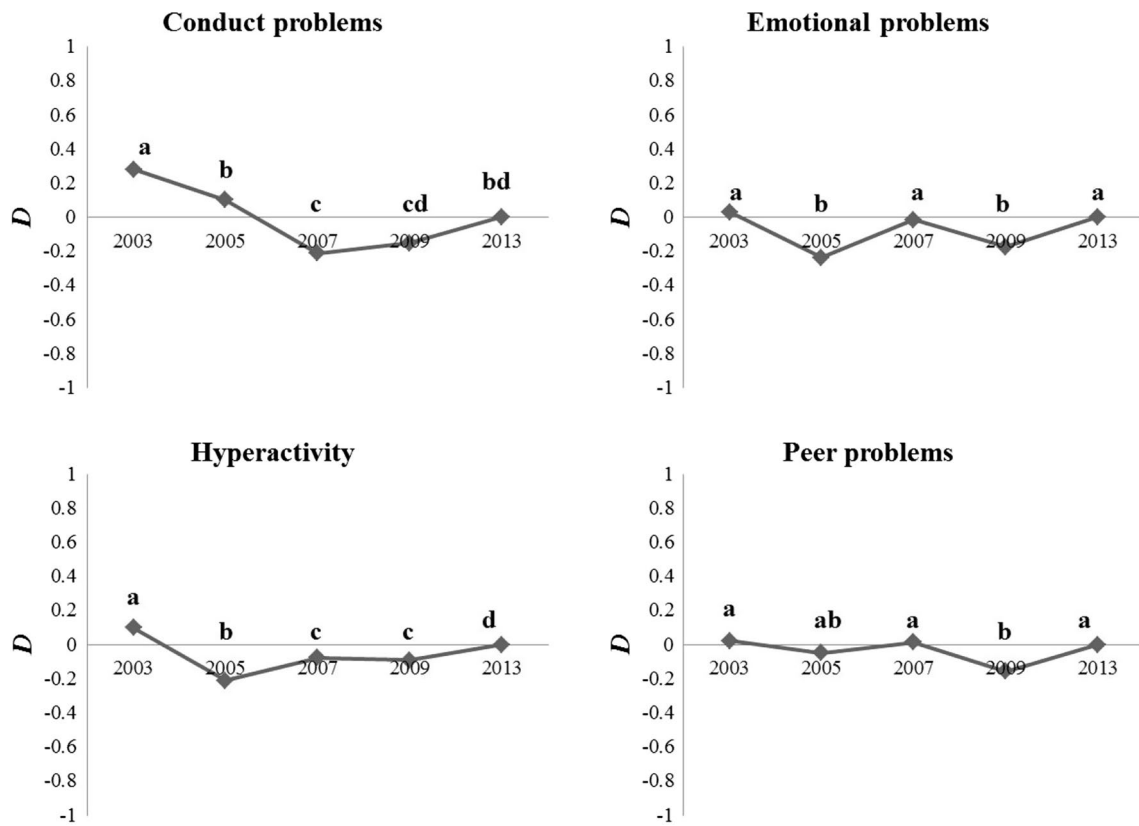


Fig. 1 Standardized latent mean differences between time points, corrected for measurement error, with 2013 as the reference category. Note: Latent means with different subscripts are significantly differ-

ent at $p < 0.05$; using other time points as the reference category similar results were found

acceptable fit to the data (see Table 2, model 3): CFI values were just below 0.900 and RMSEA values were far below 0.080. CFI values just below 0.900 were accepted, since these values could be ascribed to the low correlations between the model variables [42]. Hence, these statistics indicated that the model could be used to examine self-reported problems on all time points. Standardized factor loadings of the SDQ items and correlations between latent factors are presented in the appendix, Tables 4 and 5. All latent factors correlated significantly and positively with each other. Correlations between the peer problems and hyperactivity factors were small on all time points, whereas other latent factors showed medium to large correlations on all time points.

Next, Table 2 shows the fit indices of the models that tested measurement invariance across times and subgroups. The first model that examined configural invariance over time showed an acceptable fit, with a CFI value just below 0.900 and an RMSEA value far below 0.080. The second and third model that tested for metric and scalar invariance also yielded acceptable fits: the Chi square decreased significantly, and ΔCFI and $\Delta RMSEA$ indicated negligible changes compared to the configural model. Interestingly,

the ΔCFI value slightly exceeding the 0.010 criteria was positive, indicating that the constrained model showed a better fit to the data than the model without equality constraints. Together, these findings indicate metric and scalar invariance across time. Comparable results were found for models testing measurement invariance in gender, ethnic and educational groups. Thus, these findings showed the SDQ to be measurement invariant over time and between boys and girls, Dutch and non western adolescents, and vocational and academic educated adolescents allowing mean levels to be compared.

Trends in adolescents’ self-reported problems

Results indicated that the latent means of conduct problems [Wald $\chi^2(4) = 99.69, p < 0.001$], emotional problems [Wald $\chi^2(4) = 89.22, p < 0.001$], hyperactivity [Wald $\chi^2(4) = 72.47, p < 0.001$] and peer problems [Wald $\chi^2(4) = 10.99, p = 0.027$] were not equal across all time points. More specifically, as can be seen in Fig. 1, conduct problems decreased between 2003 and 2007, and slightly increased between 2007 and 2013. For emotional problems, similar levels of problems were found in 2003, 2007 and

Table 3 Standardized and unstandardized gender, ethnic and educational differences in problem mean levels for each time point

Subgroup variables	2003		2005		2007		2009		2013	
	B (SE)	β	B (SE)	β	B (SE)	β	B (SE)	β	B (SE)	β
Gender ^a										
Conduct problems	0.39** (0.04)	0.39	0.34** (0.05)	0.34	0.55** (0.05)	0.44	0.59** (0.05)	0.54	0.27** (0.05)	0.27
Emotional problems	-0.68** (0.04)	-0.68	-0.85** (0.05)	-0.75	-0.66** (0.03)	-0.66	-0.72** (0.04)	-0.69	-0.90** (0.03)	-0.82
Hyperactivity	-0.13** (0.03)	-0.15	-0.08* (0.03)	-0.09	-0.02 (0.03)	-0.02	0.08* (0.03)	0.08	-0.01 (0.04)	-0.01
Peer problems	0.27** (0.04)	0.28	0.16** (0.04)	0.17	0.19** (0.04)	0.19	0.23** (0.04)	0.23	0.06 (0.04)	0.06
Ethnicity ^b										
Conduct problems	0.39** (0.06)	0.35	0.23** (0.05)	0.22	0.44** (0.08)	0.31	0.33** (0.10)	0.28	0.23** (0.05)	0.22
Emotional problems	-0.09 (0.05)	-0.09	-0.15* (0.06)	-0.14	-0.10* (0.05)	-0.10	-0.21** (0.05)	-0.21	-0.11 (0.06)	-0.11
Hyperactivity	-0.34** (0.05)	-0.36	-0.40** (0.07)	-0.42	-0.34** (0.06)	-0.33	-0.35** (0.06)	-0.36	-0.26** (0.05)	-0.26
Peer problems	0.20** (0.06)	0.21	0.35** (0.06)	0.36	0.09 (0.06)	0.09	0.05 (0.08)	0.05	0.11 (0.06)	0.10
Educational level ^c										
Conduct problems	-0.42** (0.06)	-0.37	-0.52** (0.08)	-0.49	-0.52** (0.08)	-0.37	-0.63** (0.09)	-0.52	-0.46** (0.06)	-0.45
Emotional problems	-0.08 (0.05)	-0.09	-0.17* (0.07)	-0.16	-0.15** (0.04)	-0.16	-0.08 (0.06)	-0.08	-0.14* (0.07)	-0.14
Hyperactivity	-0.11* (0.04)	-0.11	-0.33** (0.06)	-0.34	-0.21** (0.04)	-0.20	-0.21** (0.04)	-0.21	-0.27** (0.06)	-0.27
Peer problems	-0.49** (0.06)	-0.50	-0.44** (0.07)	-0.45	-0.39** (0.06)	-0.39	-0.44** (0.07)	-0.43	-0.44** (0.05)	-0.43

* $p < 0.05$, ** $p < 0.001$

^a Gender (0 = girls, 1 = boys)

^b Ethnicity (0 = Dutch, 1 = non western)

^c Educational level (0 = vocational, 1 = academic)

2013, while slightly lower mean levels were revealed in 2005 and 2009. Hyperactivity problem levels also declined between 2003 and 2005, and increased again between 2005 and 2013. The most stable pattern was found for peer problems, only in 2009 problem levels were lower than most of the other years. Although these significant mean differences indicate some changes in problem levels over time, it is important to note that mean differences were only negligible ($D < 0.20$) to small ($D < 0.50$) in size.

Trends in subgroup differences

To explore whether it was necessary to control for any possibly confounding relationships between the subgroup variables, correlations between the subgroup variables were examined for each time point. No significant ($p > 0.05$) and negligible correlations between gender and ethnicity ($r_{\text{range}} = < -0.01$ to -0.05) and gender and educational level ($r_{\text{range}} = < -0.01$ to 0.08) were found. However, significant ($p < 0.05$) and small to medium correlations were found between ethnicity and educational level ($r_{\text{range}} = -0.20$ to -0.38). On each time point, non western

adolescents were less likely to attend academic educational tracks. To control for this interrelatedness, ethnicity and educational level were analyzed together in one multigroup MIMIC model. Both multigroup MIMIC models showed an acceptable model fit (see Table 2, model 13 and 14).

Gender differences in self-reported problems over time

For all problems, results indicated that mean differences between boys and girls changed across all five time points [conduct problems: Wald $\chi^2(4) = 31.71$, $p < 0.001$; emotional problems: Wald $\chi^2(4) = 47.32$, $p < 0.001$; hyperactivity: Wald $\chi^2(4) = 25.55$, $p < 0.001$; peer problems: Wald $\chi^2(4) = 18.46$, $p = 0.001$]. As can be seen in Table 3, for conduct problems gender differences were relatively strong in 2007 and 2009, whereas gender differences in emotional problems were most pronounced in 2005 and 2013. For hyperactivity, gender differences were most evident in 2003. Gender differences in peer problems disappeared in 2013. Although these findings indicated some changing gender differences in problems, in general, the significance and substantiality of these gender differences were highly

comparable between time points (see Table 3). On all time points, boys reported more conduct problems and less emotional problems than girls. In addition, on all time points except for 2013 boys reported more peer problems. Moreover, the size of gender differences in hyperactivity was not substantial on all time points.

Ethnic differences in self-reported problems over time

Comparable ethnic differences were found on all time points for conduct problems [Wald $\chi^2(4) = 8.95$, $p = 0.062$], emotional problems [Wald $\chi^2(4) = 4.02$, $p = 0.404$] and hyperactivity [Wald $\chi^2(4) = 2.87$, $p = 0.581$], indicating stable differences between Dutch and non western adolescents in these problems between 2003 and 2013. Indeed, Table 3 shows that non western adolescents were found to report more conduct problems and lower levels of hyperactivity compared to Dutch adolescents on all time points. In addition, on all time points, differences in emotional problems were negligible or just small in size. In contrast, ethnic differences in peer problems changed to some extent over time [Wald $\chi^2(4) = 16.20$, $p = 0.003$]. As can be seen in Table 3, ethnic differences were small in 2003 and 2005 with non western adolescents reporting more peer problems whereas no ethnic differences existed in 2007 and onwards.

Educational differences in self-reported problems over time

Educational level difference in conduct problems [Wald $\chi^2(4) = 5.15$, $p = 0.273$], emotional problems [Wald $\chi^2(4) = 2.27$, $p = 0.687$], and peer problems [Wald $\chi^2(4) = 1.33$, $p = 0.856$] were found to be equal across all time points, indicating that differences between vocational and academic educated adolescents in these problems were stable between 2003 and 2013. Academically educated adolescents reported less conduct problems and less peer problems compared to vocationally educated adolescents on each time point. For emotional problems, differences reached no substantial size on all time points. In contrast, educational differences in hyperactivity [Wald $\chi^2(4) = 11.38$, $p = 0.023$] changed somewhat over time. Differences between adolescents following vocational and academic educational levels were negligible in 2003, most pronounced in 2005 and small in 2007 and 2013 (see Table 3).

Discussion

It has been widely acknowledged that social, cultural, economic and governmental changes may affect adolescents' development [5]. However, whilst not ignoring the

negligible to small variations in problem mean levels over time in this study, no consistent increases or decreases in emotional and behavioral problems were found for Dutch adolescents during the last decade. Only for conduct problems, variations over time were somewhat more substantial, with the decline in conduct problems being just short of medium between 2003 and 2007. Hence, the results of the present study indicate that emotional and behavioral problem levels of Dutch adolescents were rather stable during the last decade. In line with this notion of stability, gender, ethnic and educational differences in self-reported problems on each time point were relatively stable as well, indicative of persistent mental health inequalities.

The findings suggest that variations in emotional and behavioral problems between 2003 and 2013 should not be considered as consistent changes but rather as temporary changes within this period of time. The most notable result in this respect was the small decline in conduct problems between 2003 and 2007 and the subsequent small rise in conduct problem levels afterwards. These temporary changes in self-reported problems might be important for the interpretation of the inconsistent findings in previous research on time trends. More specifically, the inconsistent findings might be the result of the fact that former studies predominantly used two measurement moments, and these two moments may not represent a consistent time trend in certain problems. To exemplify, if the present study would have examined emotional problems between 2003 and 2009 by means of two time points, results would have indicated a decrease in these problems, whereas a comparison of 2003 and 2013 would have indicated stable levels of emotional problems. Obviously, other reasons may be formulated to explain the differential results between studies as well. For instance, although it has been argued that societal changes of recent decades have taken place throughout many Western countries [28], many societal changes might have been culture and country specific and may as such be partly responsible for differences in time trends between studies conducted in different countries. In addition, different studies have covered different time spans, which may also explain the variation in results between studies.

The rather stable levels of emotional and behavioral problems of Dutch adolescents throughout the last decade may be explained by considering the impact of societal changes on adolescents' proximal environments. Whereas there is evidence for the occurrence of numerous societal changes during the last decade [11, 12, 24, 28], it can be questioned whether these changes were large enough to affect the proximal environment of adolescents known to more directly affect adolescents' emotional and behavioral problems (e.g., [22]). Gender differences in emotional and behavioral problems were also found to be relatively stable over time. Roughly speaking, differences in emotional

and behavioral problems between boys and girls can be explained by genetic differences (e.g., [30]) and by differences in gender role socialization processes [4, 6]. Drawing on the genetic explanation, stable gender differences were to be expected, since it is highly unlikely that considerable changes in the population gene pool occurred during the last decade. In addition, although gender role socialization processes have been found to be relevant for understanding gender differences in emotional and behavioral problems [36], gender roles may not have changed much during the last decade. More specifically, although societies seem to have become more gender egalitarian during the last decades [18], it might take more time for gender roles to change which may in turn postpone the impact of societal changes on boys' and girls' behaviors. Finally, ethnic and educational differences in emotional and behavioral problems were also rather stable over time. Especially, the first might be considered surprising, given evidence of an increasing intolerant political and public climate toward ethnic minorities in the Netherlands during the last decade (e.g., [29, 38, 39]). Although not ignoring the potential impact of these societal changes, possibly these changes hardly influence the day-to-day lives of ethnic minority adolescents.

By using five nationally representative samples separated by 2- to 4- year time intervals, this study provided a detailed description of trends in adolescents' self-reported problems. Still, several limitations of the current study should be considered. One limitation is that we were not able to test whether societal phenomena such as discrimination are able to explain the trends found in this study. Although this might be considered less relevant since the present study found rather stable problem levels over time, it is still important that future studies on time trends investigate possible underlying factors of trends in self-reported problems. A second limitation is that results were solely based on self-reports. Although self-reports are an important source of information, especially with regard to emotional problems that are less apparent to teachers and parents [27], a more comprehensive view on problem behaviours would have been obtained if parent and teacher reports were included. Third, we cannot rule out entirely that differences in sample compositions between our five samples may have affected the results. However, considering the negligibly small differences in

sample compositions between the samples, it is highly unlikely that this impact is sizable. Fourth, adolescents' willingness to report problems may have changed over time. However, in such instance, a systematic change in all self-reported problem levels would be expected. Since trends were specific for each problem behavior, it is unlikely that results were due to changes in reporting thresholds. Last, the data allowed no comparison of problem differences between specific groups of ethnic minorities. This is unfortunate, since self-reported problem levels are found to differ between ethnic minorities [32]. Future research analyzing trends for specific groups of ethnic minorities is needed to prevent a misguided generalization of findings to all ethnic minorities groups in the Netherlands.

Conclusion

The present study revealed rather stable trends in emotional and behavioral problems as well as rather stable gender, ethnic and educational differences in these trends. Hence, the most important lesson to be learnt from these findings is the stability of mental health on a population level as well as the persistency of inequalities in mental health. Among others, vocationally educated adolescents reported higher levels of hyperactivity and more conduct and peer problems compared to their academically educated peers, ethnic minorities reported more conduct problems and ethnic majority adolescents reported higher levels of hyperactivity on all time points. To change these persistent mental health inequalities, intervention and prevention efforts should be tailored to the specific needs of these more vulnerable groups of adolescents. Moreover, internationally comparative research using multiple measurement moments is needed to examine whether these stable differences in emotional and behavioral problems hold over extended periods of time in different countries.

Conflict of interest None.

Appendix

See Appendix Tables 4 and 5.

Table 4 Standardized factor loadings of the SDQ items in the multigroup four factor model

Latent factors	β (2003)	β (2005)	β (2007)	β (2009)	β (2013)	Range (R^2)
Conduct problems						
Tempers	0.707	0.706	0.728	0.730	0.728	0.498–0.533
Obedient ^a	–0.195	–0.190	–0.287	–0.229	–0.204	0.036–0.082
Fights or bullies	0.538	0.531	0.588	0.586	0.510	0.260–0.346
Lies or cheats	0.696	0.694	0.686	0.689	0.675	0.457–0.482
Steals	0.502	0.468	0.486	0.500	0.510	0.219–0.260
Emotional problems						
Somatic symptoms	0.530	0.575	0.508	0.543	0.567	0.258–0.330
Worries	0.618	0.668	0.633	0.661	0.647	0.382–0.446
Unhappy	0.844	0.808	0.832	0.834	0.880	0.653–0.775
Nervous in new situations	0.604	0.640	0.596	0.634	0.644	0.355–0.415
Many fears	0.663	0.673	0.648	0.688	0.676	0.420–0.473
Hyperactivity						
Restless	0.535	0.645	0.632	0.604	0.655	0.286–0.429
Fidgety	0.574	0.624	0.611	0.598	0.613	0.329–0.389
Easily distracted	0.791	0.759	0.836	0.825	0.832	0.576–0.700
Thinks before acting ^a	–0.479	–0.459	–0.505	–0.480	–0.498	0.210–0.255
Good attention ^a	–0.577	–0.560	–0.659	–0.634	–0.632	0.314–0.434
Peer problems						
Solitary	0.463	0.460	0.461	0.505	0.471	0.211–0.255
Has good friend ^a	–0.207	–0.192	–0.235	–0.262	–0.262	0.037–0.069
Generally liked ^a	–0.454	–0.413	–0.506	–0.519	–0.498	0.170–0.270
Picked on or bullied	0.774	0.828	0.771	0.810	0.832	0.594–0.692
Better with adults than with children	0.471	0.493	0.459	0.488	0.514	0.211–0.269

All loadings were significant at $p < 0.001$

^a Positively phrased items with higher scores indicating less problems

Table 5 Correlations between the latent factors on each time point

Latent factors	2003	2005	2007	2009	2013
Emotional problems with Conduct problems	0.423	0.468	0.454	0.390	0.507
Hyperactivity with Conduct problems	0.569	0.635	0.650	0.588	0.669
Hyperactivity with Emotional problems	0.418	0.465	0.412	0.368	0.405
Peer problems with Conduct problems	0.523	0.569	0.528	0.505	0.635
Peer problems with Emotional problems	0.569	0.589	0.646	0.562	0.602
Peer problems with Hyperactivity	0.138	0.194	0.203	0.163	0.225

All correlations were significant at $p < 0.001$

References

- Achenbach TM, Becker A, Döpfner M, Heiervang E, Roessner V, Steinhausen HC, Rothenberger A (2008) Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions. *J Child Psychol Psyc* 49:251–275
- Achenbach TM, Dumenci L, Rescorla LA (2002) Ten-year comparisons of problems and competencies for national samples of youth self, parent, and teacher reports. *J Emot Behav Disord* 10:194–203
- Achenbach TM, Dumenci L, Rescorla LA (2003) Are American children's problems still getting worse? A 23-year comparison. *J Abnorm Child Psych* 31:1–11
- Bem SL (1981) Gender schema theory: a cognitive account of sex typing. *Psychol Rev* 88:354–364
- Bronfenbrenner U (1979) *The ecology of human development: experiments by nature and design*. Harvard University Press, Cambridge
- Bussey K, Bandura A (1992) Self-regulatory mechanisms governing gender development. *Child Dev* 63:1236–1250
- Byrne BM (2012) *Structural equation modeling with Mplus: basic concepts, applications, and programming*. Routledge, New York
- Chen FF (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Model* 14:464–504
- Choi J, Fan W, Hancock GR (2009) A note on confidence intervals for two-group latent mean effect size measures. *Multivar Behav Res* 44:396–406

10. Cohen J (1988) Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates, Hillsdale
11. Collishaw S, Gardner F, Maughan B, Scott J, Pickles A (2012) Do historical changes in parent–child relationships explain increases in youth conduct problems? *J Abnorm Child Psych* 40:119–132
12. Collishaw S, Goodman R, Pickles A, Maughan B (2007) Modelling the contribution of changes in family life to time trends in adolescent conduct problems. *Soc Sci Med* 65:2576–2587
13. Collishaw S, Maughan B, Goodman R, Pickles A (2004) Time trends in adolescent mental health. *J Child Psychol Psych* 45:1350–1362
14. Collishaw S, Maughan B, Natarajan L, Pickles A (2010) Trends in adolescent emotional problems in England: a comparison of two national cohorts twenty years apart. *J Child Psychol Psychiatry* 51:885–894
15. De Looze M, van Dorsselaer S, De Roos S, Verdurmen J, Stevens G, Gommans R, Vollebergh W (2013) HBSC 2013: Gezondheid, welzijn en opvoeding van jongeren in Nederland [Health, well-being and upbringing of youth in the Netherlands]. Netherlands Institute of Mental Health and Addiction, Utrecht
16. van Dorsselaer S, de Looze M, Vermeulen-Smit E, de Roos S, Verdurmen J, ter Bogt T, Vollebergh W (2010) HBSC 2009: Gezondheid en welzijn van jongeren in Nederland [Health, well-being and upbringing of youth in the Netherlands]. Netherlands Institute of Mental Health and Addiction, Utrecht
17. van Dorsselaer S, Zeijl E, van den Eeckhout S, Ter Bogt T, Vollebergh WAM (2007) HBSC 2005: Gezondheid, welzijn en opvoeding van jongeren in Nederland [Health, well-being and upbringing of youth in the Netherlands]. Netherlands Institute of Mental Health and Addiction, Utrecht
18. England P (2010) The gender revolution uneven and stalled. *Gender Soc* 24:149–166
19. Fichter MM, Xepapadakis F, Quadflieg N, Georgopoulou E, Fthenakis WE (2004) A comparative study of psychopathology in Greek adolescents in Germany and in Greece in 1980 and 1998—18 years apart. *Eur Arch Psychiatry Clin Neurosci* 254:27–35
20. Goodman R (1997) The Strengths and Difficulties Questionnaire: a research note. *J Child Psychol Psych* 38:581–586
21. Goodman A, Lamping DL, Ploubidis GB (2010) When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. *J Abnorm Child Psych* 38:1179–1191
22. Hill J (2002) Biological, psychological and social processes in the conduct disorders. *J Child Psychol Psych* 43:133–164
23. Maughan B, Collishaw S, Meltzer H, Goodman R (2008) Recent trends in UK child and adolescent mental health. *Soc Psychiatry Psychiatr Epidemiol* 43:305–310
24. Maughan B, Iervolino AC, Collishaw S (2005) Time trends in child and adolescent mental disorders. *Curr Opin Psychiatry* 18:381–385
25. Monshouwer K, van Dorsselaer S, Gorter A, Verdurmen J, Vollebergh W (2004) Jeugd en riskant gedrag 2003 [Youth and risky behavior]. Netherlands Institute of Mental Health and Addiction, Utrecht
26. Monshouwer K, Verdurmen J, van Dorsselaer S, Gorter A, Vollebergh W (2008) Jeugd en riskant gedrag 2007 [Youth and risky behavior]. Netherlands Institute of Mental Health and Addiction, Utrecht
27. Rescorla LA, Ginzburg S, Achenbach TM et al (2013) Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *J Clin Child Adolesc* 42:262–273
28. Rutter M, Smith DJ (1995) Psychosocial disorders in young people: time trends and their causes. Wiley, Chichester
29. Shadid WA (2006) Public debates over Islam and the awareness of Muslim identity in the Netherlands. *Eur Educ* 38:10–22
30. Silberg J, Pickles A, Rutter M et al (1999) The influence of genetic factors and life stress on depression among adolescent girls. *Arch Gen Psychiatry* 56:225–232
31. Sourander A, Koskelainen M, Niemelä S, Rihko M, Ristkari T, Lindroos J (2012) Changes in adolescents mental health and use of alcohol and tobacco: a 10-year time-trend study of Finnish adolescents. *Eur Child Adolesc Psychiatry* 21:665–671
32. Stevens GWJM, Pels T, Bengi-Arslan L, Verhulst FC, Vollebergh WAM, Crijnen AAM (2003) Parent, teacher and self-reported problem behavior in the Netherlands. *Soc Psychiatry Psychiatr Epidemiol* 38:576–585
33. Sweeting H, Young P, West P (2009) GHQ increases among Scottish 15 year olds 1987–2006. *Soc Psychiatry Psychiatr Epidemiol* 44:579–586
34. Tick NT, van der Ende J, Verhulst FC (2007) Twenty-year trends in emotional and behavioral problems in Dutch children in a changing society. *Acta Psychiatr Scand* 116:473–482
35. Tick NT, van der Ende J, Verhulst FC (2008) Ten-year trends in self-reported emotional and behavioral problems of Dutch adolescents. *Soc Psychiatry Psychiatr Epidemiol* 43:349–355
36. Torsheim T, Ravens-Sieberer U, Hetland J, Välimaa R, Danielson M, Overpeck M (2006) Cross-national variation of gender differences in adolescent subjective health in Europe and North America. *Soc Sci Med* 62:815–827
37. Twenge JM, Gentile B, DeWall CN, Ma D, Lacefield K, Schurtz DR (2010) Birth cohort increases in psychopathology among young Americans, 1938–2007: a cross-temporal meta-analysis of the MMPI. *Clin Psychol Rev* 30:145–154
38. Van der Veer P (2006) Pim Fortuyn, Theo van Gogh, and the politics of tolerance in the Netherlands. *Public Cult* 18:111–124
39. Vasta E (2007) From ethnic minorities to ethnic majority policy: multiculturalism and the shift to assimilationism in the Netherlands. *Ethnic Racial Stud* 30:713–740
40. Verhulst FC, van der Ende J, Rietbergen A (1997) Ten-year time trends of psychopathology in Dutch children and adolescents: no evidence for strong trends. *Acta Psychiatr Scand* 96:7–13
41. Wångby M, Magnusson D, Stattin H (2005) Time trends in the adjustment of Swedish teenage girls: a 26-year comparison of 15-year-olds. *Scand J Psychol* 46:145–156
42. Wang J, Wang X (2012) Structural equation modeling: applications using Mplus. Wiley, Chichester
43. West P, Sweeting H (2003) Fifteen, female and stressed: changing patterns of psychological distress over time. *J Child Psychol Psychiatr* 44:399–411
44. van Widenfelt BM, Goedhart AW, Treffers PDA, Goodman R (2003) Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *Eur Child Adolesc Psychiatry* 10:281–289