OHTE GENERAL REPORT ON THE STATE OF HISTORY TEACHING IN EUROPE 2023

VOLUME 3
TECHNICAL APPENDIX

OBSERVATORY ON HISTORY TEACHING IN EUROPE

COUNCIL OF EUROPE
75
1949 - 2024
CONSEIL DE L'EUROPE

All other correspondence concerning this document should be addressed to the Observatory on History Teaching in Europe

F-67075 Strasbourg Cedex, France
E-mail: ohte@coe.int

Cover design and layout: ZN, Belgium

Cover photos: Europeana, The Auschwitz Museum, Bibliothèque nationale de France, @Lefteris on Canva, Canva pro, Adobe Stock

Council of Europe Publishing
F-67075 Strasbourg Cedex
http://book.coe.int

ISBN: 978-92-871-9494-7
ISSN: 3007-908X

Printed by ZN, Belgium

OHTE GENERAL REPORT ON THE STATE OF HISTORY TEACHING IN EUROPE

2023

VOLUME 3
TECHNICAL APPENDIX

COUNCIL OF EUROPE

CONTENTS

ITEM 1

RESEARCH INSTRUMENTS

Item 1

Research instruments

EDUCATION AUTHORITIES' SURVEY (EAS)

https://rm.coe.int/2023-07-appendix-education-authorities-ohte-regular-report-questionnai/native/1680abf18a

TEACHERS' AND EDUCATORS' SURVEY (TES)

https://rm.coe.int/2023-07-appendix-teachers-and-educators-ohte-regular-report-questionna/native/1680abf18c

OVERVIEW OF HISTORY COURSES TEMPLATE

https://rm.coe.int/2023-07-appendix-education-authorities-ohte-regular-report-overview-of/native/1680abf18b

Table 1 – List of educator focus groups (EFGs)

	Educator focus groups	Date	Place	Number of participants
1st round	Focus Group 1	2 December 2022	Strasbourg	9
	Focus Group 2	25 January 2023	Online	4
	Focus Group 3	26 January 2023	Online	4
	Focus Group 4	1 February 2023	Online	3
	Focus Group 5	2 February 2023	Online	6
2nd round	Focus Group 6	8 March 2023	Brussels	5
	Focus Group 7	8 March 2023	Brussels	4
	Focus Group 8	9 March 2023	Brussels	5
3rd round	Focus Group 9	20 April 2023	Vilnius	2
	Focus Group 10	22 April 2023	Vilnius	2
	Focus Group 11	22 April 2023	Vilnius	5

ITEM 2

ANALYSIS OF THE RELIABILITY AND VALIDITY IN THE TES

Item 2

Analysis of the reliability and validity in the TES

DATA ANALYSIS

The analysis strategy consisted of three stages. In the first stage, the database was cleaned up and the data were organised for subsequent analysis. First, participants who had not completed at least one of the subscales in addition to the initial subscale of identification data were eliminated because they did not provide information of relevance to the study. Responses to certain questions relating to category (e.g., gender) that had been left open, giving rise to multiple categories (e.g., female, feminine, mujer, women, etc.; Greek, greek, Spanish, español, etc.), were standardised. In addition, typographical errors were corrected (e.g., in question TII.5, "How effectively do you think the history curriculum of your country addresses diversity?", which should be answered on a scale from 1 to 10, we came across data out of the range [e.g., "0"], since the question required an open-ended response rather than the selection of an item on a numbered scale). Errors of this type should be avoided when setting the questions, with participants being given a response scale. Lastly, the names of the variables were assigned to the database headings.

In the second stage, a descriptive analysis was made of the responses to each section of the questionnaire, and the central tendency and dispersion measures were analysed.

Finally, in the third stage, the reliability indices (Cronbach's alpha and McDonald's omega) and multivariate outliers, via the D2 Mahalanobis distances and Guttman errors, were analysed. Mokken scale analysis was used to assess whether the scoring of the different items in each subscale reflected the same latent variable. The Mokken scale is an item response model in biometrics that is usually employed to assess measurement scales in psychology (Molenaar and Sjitsma 1984). Item scalability was assessed by means of Loevinger's homogeneity coefficient (H). The homogeneity coefficients (H) obtained make it possible to assess the unidimensional nature of the subscales. The cut-off values used in previous studies were considered (Molenaar and Sjitsma 1984; Stochl et al. 2012). All H values should exceed 0.3 on a unidimensional scale. Values between 0.3 and 0.4 indicate low accuracy, those between 0.4 and 0.5 indicate average accuracy and those in excess of 0.5 indicate high accuracy (Stochl et al. 2012). The automatic item selection procedure (AISP) was then used to divide the whole range of items into unidimensional scales (Ark 2007). In addition, the cases in which those questioned selected response options that were inconsistent with the expected general pattern (Guttman errors) were analysed. The basic idea is to compare the quantity of errors observed with the quantity of errors expected under the marginal independence model (Loevinger 1948; Mokken 1971). Software R version 4.0.4 (2021-02-15) was used for the data analysis.

RESULTS

The results of the following are presented below:

a Descriptive analysis;

b Reliability analysis;

c Mokken scalability – homogeneity coefficients and automatic item selection procedure;

d Multivariate outliers;

e Evidence relating to construct validity, for each scale included in the questionnaire.

SUBSCALE II – History curricula

a Descriptive analysis

Table 2.1 shows the results of responses to the following items: 49 ("How rigid is the curriculum structure and its requirements, and how much room for discretion is there for you to organise your teaching?"); 50 ("How manageable is the amount of content that you have to cover according to the curricula?"); and 55 ("How effectively do you think the history curriculum addresses diversity?").

Table 2.1 – Subscale II descriptives

Item	Variable	1	2	3	4	5	% 1 to 2	% 4 to 5	Missing
Flexibility	i49	356	953	2 019	1 360	787	16.65	27.3	2 392
Density	i50	459	1 173	1 504	1 345	929	20.75	28.9	2 457
Diversity	i55	252	985	1 766	1 289	798	15.73	26.54	2 777

Item	n	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
i49	5475	3.23	1.10	3	1	5	4	-0.11	-0.58	0.01
i50	5410	3.21	1.20	3	1	5	4	-0.10	-0.93	0.02
i55	5090	3.27	1.09	3	1	5	4	-0.07	-0.70	0.02

b Reliability

With respect to the reliability of the three items, values higher than 0.7 were obtained by means of both Cronbach's ordinal alpha ($\alpha = 0.75$) and McDonald's ordinal omega ($\omega = 0.75$) (McDonald 2013; Revelle and Zinbarg 2009). Values higher than 0.7 indicate good reliability (Kline 1999).

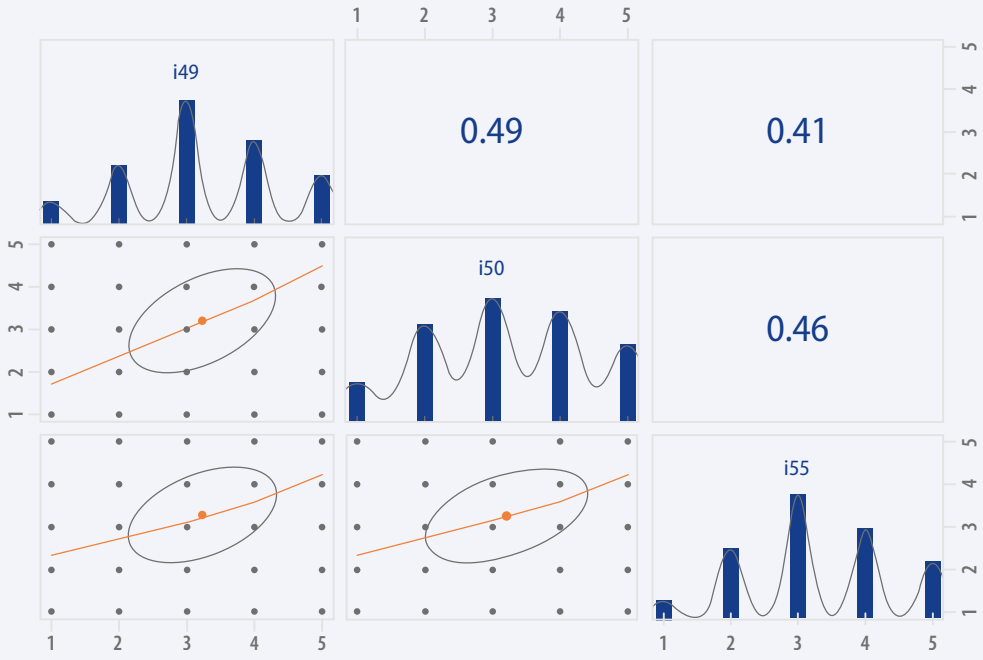
Table 2.2 presents the results of the reliability analysis and the item–total correlations of the scale. It can be seen that all item–total correlations were greater than 0.3 and that the elimination of any item does not substantially improve the reliability of the subscale.

Table 2.2 – Subscale II reliability analysis (TII)

	Alpha if an item is dropped	Item-total correlation
i49	0.67	0.57
i50	0.62	0.61
i55	0.70	0.54

The distributions of the three variables composing the subscale and the correlations between them are shown in **Figure 2.1**.

Figure 2.1 – Frequency distributions and correlations between the variables of Subscale II (TII)



c Mokken scale analysis

With respect to item homogeneity, the homogeneity coefficients (H) are examined for the set of items (for each item, item pair and the general scale). The general scalability coefficient obtained for the three items was H = 0 .469 (SE = 0.01). The scalability of the item pairs fell between Hij = 0.513 (SE = 0.013) for item pair 49-50, Hij = 0.418 (SE = 0.014) for item pair 49-55, and Hij = 0.474 (SE = 0.013) for item pair 50-55. Multidimensionality indices were therefore not identified and the items are scalable to H ≥0.30, which indicates average accuracy (Stochl et al. 2012).

The automated item selection procedure was then carried out at increasing homogeneity

threshold levels to examine dimensionality. If all items are designated as belonging to dimension 1, this indicates that the scale is unidimensional within that homogeneity threshold (indicated in the column headings, from 0.1 to 0.5). **Table 2.3** shows the results of the AISP, demonstrating that the three items together can be considered to be unidimensional, with a homogeneity threshold of H ≥ 0.3.

Table 2.3 – MSA-AISP for increasing H thresholds *(t)* – Subscale II

Item	*t* = 0.10	*t* = 0.15	*t* = 0.20	*t* = 0.30	*t* = 0.35	*t* = 0.40	*t* = 0.45	*t* = 0.50
i49	1	1	1	1	1	1	1	1
i50	1	1	1	1	1	1	1	1
i55	1	1	1	1	1	1	0	0

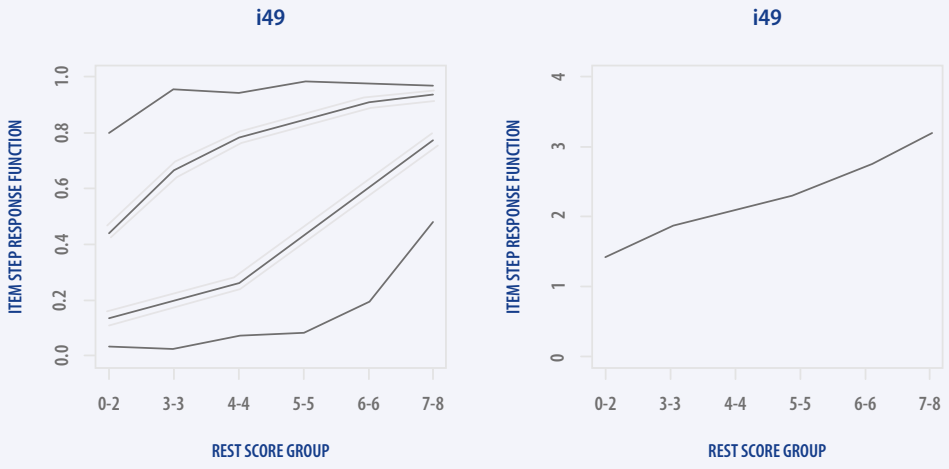
With respect to the monotonicity assumption, **Table 2.4** shows the results of the analysis, which indicate that there are no significant violations (#zsig) nor any insignificant violations (#vi) of monotonicity for any of the items in Subscale II. That is, all items appear to discriminate clearly between those questioned with high levels in the construct and those with lower levels.

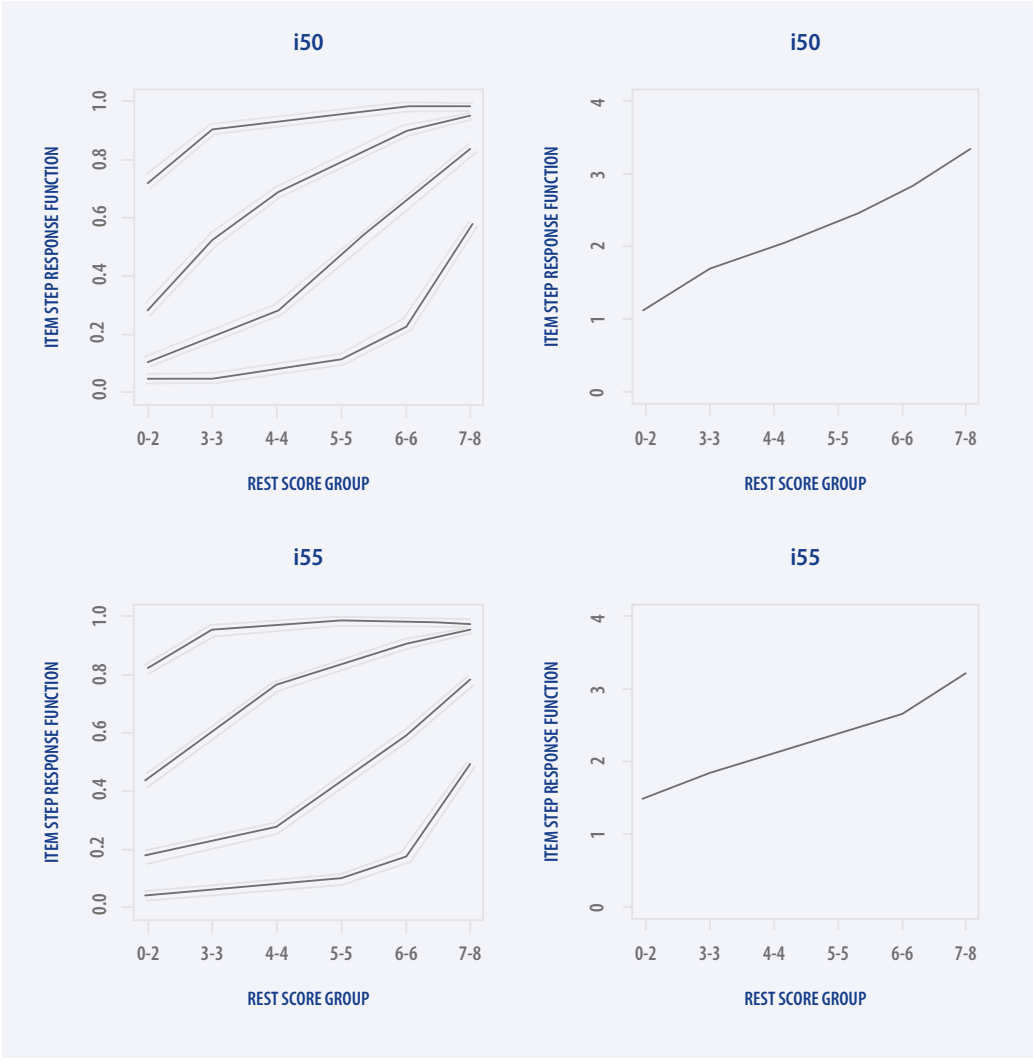
Table 2.4 – MSA – Monotonicity – Subscale II

Item	H	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i49	0.47	60	0	0	0	0	0	0	0	0
i50	0.49	60	0	0	0	0	0	0	0	0
i55	0.45	60	0	0	0	0	0	0	0	0

Figure 2.2 shows monotonically increasing item step response functions (ISRF).

Figure 2.2 – Step response function (ISRF) – Subscale II

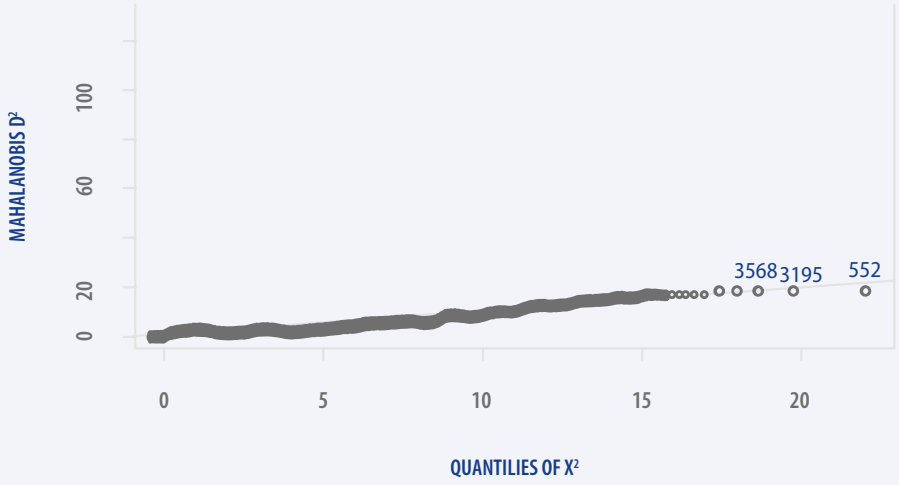




d Multivariate outliers

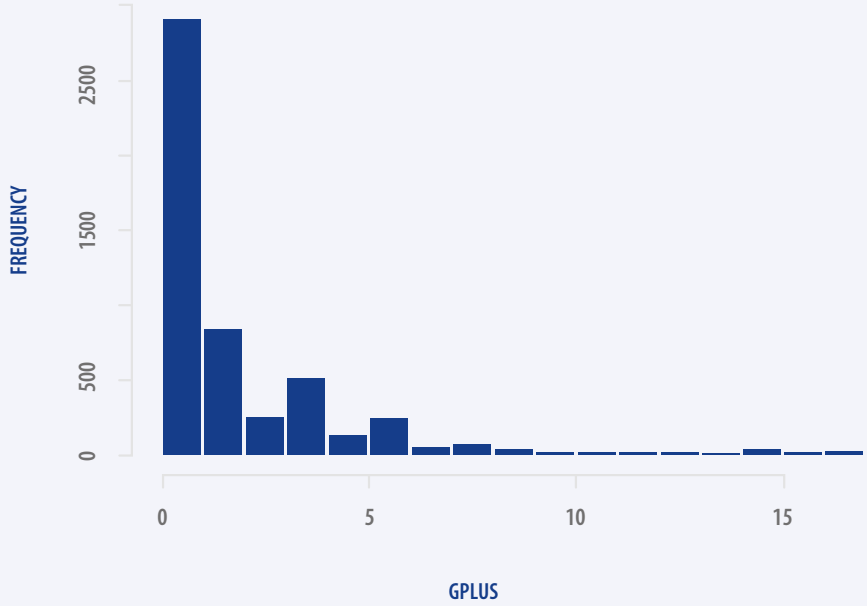
Multivariate outliers were then analysed by means of the Mahalanobis D^2 distances. The respective results are shown in **Figure 2.3**. As can be seen, the D^2 distances are shown on the y-axis while the chi-squared quantiles are shown on the x-axis. Five of the D^2 distance values were significant at confidence level $\alpha = 0.001$ (Hair et al. 2019). These were cases 552, 3 195, 3 568, 6 980 and 7 040, the maximum value of D^2 being 18.06.

Figure 2.3 – Q–Q plot of Mahalanobis D^2 vs quantiles of χ^2 – Subscale II



The number of Guttman errors for each observation was then calculated to identify atypical response patterns. The error average was 1.96 (SD = 2.81), according to the criterion proposed by Zijlstra et al. (2007) and Hubert and Vandervieren (2008) for asymmetric distributions, and the critical value was 5. Thus 479 cases with atypical response patterns were identified. **Figure 2.4** shows the distribution of Guttman errors.

Figure 2.4 – Distribution of Guttman errors – Subscale II



e Evidence of validity

To verify data adequacy for factor analysis, the Kaiser Meyer Olkin (KMO) (Kaiser 1970) and Bartlett sphericity tests were used. The KMO statistic is a measure of the adequacy of data for factor analysis, that is, indicating whether the data are adequate for carrying out a factor analysis of the relationship linking the correlations between the items and the partial correlations, that is, this test seeks to respond to the question "Are the data adequate for factor analysis?" Kaiser and Rice (1974) suggest that KMO values below 0.5 are unacceptable for such analysis, while values above 0.6 are considered to be mediocre, above 0.7 acceptable, above 0.8 commendable and above 0.9 excellent. The result of the KMO test indicates that the data are adequate for factor analysis (KMO = 0.69). In addition to the KMO measure for the complete test, it is possible to verify the sample adequacy measures for each test indicator by means of the individual measure of sampling adequacy (MSA) (Kaiser 1970; Kaiser and Rice 1974; Lorenzo-Seva and Ferrando 2021). Once again, MSA values close to 1 will indicate that each item (considered individually in this case) is adequate for submission to factor analysis, while items with MSA values below 0.5 should be omitted from factor analysis (Lorenzo-Seva and Ferrando 2021). In this case, the three questionnaire items obtained MSA values in excess of 0.6 (i49 = 0.69, i50 = 0.66, i55 = 0.72). In light of those results, it appears appropriate to include an additional item of some kind in the questionnaire on the curriculum. The internal structure is shown in **Figure 2.5**.

For its part, Bartlett's sphericity test checks the null hypothesis that the correlations matrix is an identity matrix (a matrix in which the elements outside the diagonal are all 0, such that there would not be any correlation between the variables). The results of the test should be significant ($p < 0\ .01$). In this case, once again, the results obtained suggested that the data are adequate for submission to factor analysis (Bartlett's sphericity test, $\chi^2\ (3) = 68.1; p < 0.001$).

Figure 2.5 – Measurement module – Subscale II

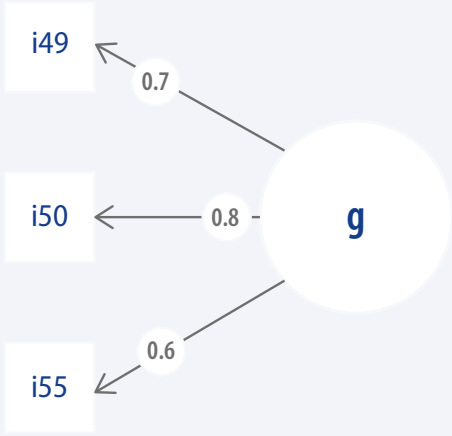
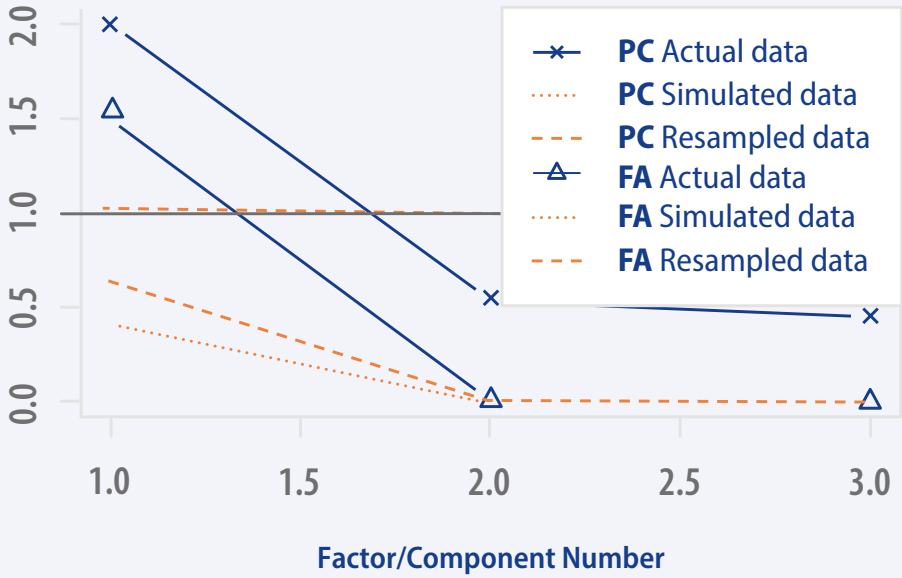


Figure 2.6 presents the sedimentation graph with the results of exploratory factor analysis, which suggests the presence of one factor. The presence of one factor is supported by 10 methods out of 18 (55.56%) (optimal co-ordinates, acceleration factor, parallel analysis, Kaiser criterion, Scree (SE), Exploratory Graph Analysis (EGA) (glasso), Exploratory Graph Analysis (Triangulated Maximally Filtered Graph) (EGA (TMFG)), Very Simple Structure (VSS) complexity 1, Velicer's Minimum Average Partial (MAP), Tucker-Lewis index (TLI)).

Figure 2.6 – Sedimentation graph – Subscale II



SUBSCALE III – History textbooks and educational resources

Subscale III.I: use of educational resources

a Descriptive analysis

The results of the descriptive analysis of the 17 items relating to the frequency of use of different educational resources (Subscale III.I, items 56.1 to 56.17) are shown in **Table 2.5**.

Table 2.5 – Frequency of use of different resources – Subscale III.I (items 56.1 to 56.17)

Item	Item label	1	2	3	4	5	% 1 to 2	% 4 to 5	Missing
Apps for smartphones and tablets with historical content	i56.1	1 164	1 067	1 405	978	359	44.86	26.89	2 894
Artefacts (e.g., paintings, architecture, sculptures, contemporary art)	i56.2	406	1 245	1 644	1 249	429	33.2	33.74	2 894

Item	Item label	1	2	3	4	5	% 1 to 2	% 4 to 5	Missing
Cinema and documentaries with historical themes	i56.3	329	1 001	1 669	1 613	361	26.74	39.69	2 894
Historical bibliography	i56.4	478	1 263	1 656	1 151	425	35.01	31.69	2 894
Historical novels, comics and children's literature	i56.5	795	1 561	1 540	863	214	47.38	21.66	2 894
Local and regional festivals and traditions related to historical events	i56.6	1 122	1 623	1 325	723	180	55.2	18.16	2 894
Local cultural heritage (e.g., costumes, food traditions, celebrations)	i56.7	284	1 192	1 634	1 417	446	29.68	37.46	2 894
Museums and other places of heritage interpretation	i56.8	378	1 353	1 655	1 241	346	34.81	31.91	2 894
History textbooks	i56.9	98	258	496	897	3 224	7.16	82.87	2 894
Oral sources (interviews with grandparents, relatives, neighbours, etc.)	i56.10	692	1 478	1 518	904	381	43.64	25.84	2 894
Primary documentary sources	i56.11	312	973	1 502	1 383	803	25.84	43.96	2 894
Printed or digital press (newspapers and magazines)	i56.12	256	997	1 639	1 495	586	25.2	41.85	2 894
Reports on historical topics in popular magazines	i56.13	816	1 511	1 578	867	201	46.79	21.48	2 894
Search engines and websites with historical content not necessarily validated by the education authorities	i56.14	549	943	1 412	1 476	593	30	41.6	2 894
Teacher notes	i56.15	274	606	1 073	1 423	1 597	17.7	60.73	2 894
Video games	i56.16	2 420	1 290	792	368	103	74.6	9.47	2 894
Websites and databases with historical content approved by the education authorities	i56.17	217	781	1 448	1 736	791	20.07	50.81	2 894

Item	Vars	n	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
Apps for smartphones and tablets with historical content	i56.1	4 973	2.66	1.23	3	1	5	4	0.15	-1	0.02
Artefacts (e.g., paintings, architecture, sculptures, contemporary art)	i56.2	4 973	3.01	1.08	3	1	5	4	0	-0.68	0.02
Cinema and documentaries with historical themes	i56.3	4 973	3.14	1.03	3	1	5	4	-0.24	-0.55	0.01
Historical bibliography	i56.4	4 973	2.96	1.1	3	1	5	4	0.04	-0.69	0.02
Historical novels, comics and children's literature	i56.5	4 973	2.63	1.08	3	1	5	4	0.22	-0.65	0.02
Local and regional festivals and traditions related to historical events	i56.6	4 973	2.44	1.1	2	1	5	4	0.38	-0.65	0.02
Local cultural heritage (e.g., costumes, food traditions, celebrations)	i56.7	4 973	3.11	1.05	3	1	5	4	-0.05	-0.67	0.01
Museums and other places of heritage interpretation	i56.8	4 973	2.96	1.05	3	1	5	4	0.04	-0.66	0.01
History textbooks	i56.9	4 973	4.39	0.99	5	1	5	4	-1.63	1.89	0.01
Oral sources (interviews with grandparents, relatives, neighbours, etc.)	i56.10	4 973	2.76	1.13	3	1	5	4	0.22	-0.7	0.02
Primary documentary sources	i56.11	4 973	3.28	1.14	3	1	5	4	-0.16	-0.78	0.02

Item	Vars	n	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
Printed or digital press (newspapers and magazines)	i56.12	4 973	3.23	1.06	3	1	5	4	-0.14	-0.63	0.02
Reports on historical topics in popular magazines	i56.13	4 973	2.62	1.07	3	1	5	4	0.19	-0.67	0.02
Search engines and websites with historical content not necessarily validated by the education authorities	i56.14	4 973	3.12	1.18	3	1	5	4	-0.21	-0.82	0.02
Teacher notes	i56.15	4 973	3.7	1.2	4	1	5	4	-0.6	-0.6	0.02
Video games	i56.16	4 973	1.88	1.06	2	1	5	4	1.03	0.21	0.01
Websites and databases with historical content approved by the education authorities	i56.17	4 973	3.42	1.07	4	1	5	4	-0.33	-0.55	0.02

b Reliability

With respect to the reliability of this set of 17 items, values higher than 0.8 were obtained by means of both Cronbach's ordinal alpha ($\alpha = 0.88$) and McDonald's omega ($\omega = 0.9$) (McDonald 2013; Revelle and Zinbarg 2009). Values higher than 0.8 are considered to be good (Kline 1999). **Table 2.6** shows the results of the reliability analysis and the item–total correlations of the scale.

It can be seen that all item–total correlations were greater than 0.3, except in the case of item 56.9 (textbooks).

Table 2.6 – Reliability analysis – Subscale III.I

Item	Alpha if an item is dropped	Item–total correlation
i56.1	0.85	0.43
i56.2	0.85	0.48
i56.3	0.85	0.49
i56.4	0.85	0.6
i56.5	0.84	0.62
i56.6	0.85	0.62
i56.7	0.85	0.65
i56.8	0.86	0.57
i56.9	0.86	0.19
i56.10	0.85	0.56
i56.11	0.85	0.50
i56.12	0.85	0.61
i56.13	0.85	0.66
i56.14	0.85	0.39
i56.15	0.85	0.34
i56.16	0.87	0.47
i56.17	0.85	0.53

c Mokken scale analysis

With respect to item homogeneity, the general scalability coefficient obtained for the 17 items was H = 0.297 (SE = 0.005). The H scalability values of all items are shown in **Table 2.7**.

Table 2.7 – Homogeneity coefficients – Subscale III.I

Item	H	SE
i56.1	0.249	(0.009)
i56.2	0.278	(0.008)
i56.3	0.284	(0.008)
i56.4	0.341	(0.008)
i56.5	0.358	(0.008)
i56.6	0.360	(0.008)
i56.7	0.366	(0.007)
i56.8	0.326	(0.008)
i56.9	0.118	(0.011)
i56.10	0.321	(0.008)
i56.11	0.288	(0.008)

Item	H	SE
i56.12	0.348	(0.007)
i56.13	0.378	(0.007)
i56.14	0.224	(0.009)
i56.15	0.196	(0.009)
i56.16	0.289	(0.010)
i56.17	0.307	(0.008)

These values once again show the possible multidimensionality of the scale. The automated item selection procedure was then carried out at increasing thresholds of homogeneity to examine dimensionality. As has been shown, if all items appear to belong to dimension 1, this indicates that the scale is unidimensional in that homogeneity threshold (indicated in the column headings, from 0.1 to 0.5). **Table 2.8** shows the results of the AISP and identifies the dimensions in the item set with a homogeneity threshold of H ≥ 0.3. Items i59.1, i56.9, i56.14 and i59.15 would remain in dimension 2, with the remainder in dimension 1. These results confirm the presence of a multidimensional structure in the scale that would have to be verified by means of exploratory factor analysis.

Table 2.8 – MSA–AISP for increasing H thresholds *(c)* – Subscale III.I

Item	*c* = 0.10	*c* = 0.15	*c* = 0.20	*c* = 0.30	*c* = 0.35	*c* = 0.40	*c* = 0.45	*c* = 0.50
i56.1	1	1	1	0	0	0	0	0
i56.2	1	1	1	1	2	0	0	0
i56.3	1	1	1	1	2	0	0	0
i56.4	1	1	1	1	1	2	0	0
i56.5	1	1	1	1	1	1	0	0
i56.6	1	1	1	1	1	1	1	1
i56.7	1	1	1	1	1	1	1	1
i56.8	1	1	1	1	1	1	1	1
i56.9	0	0	2	0	0	0	0	0
i56.10	1	1	1	1	1	1	0	0
i56.11	1	1	1	1	1	2	0	0
i56.12	1	1	1	1	1	1	2	2
i56.13	1	1	1	1	1	1	2	2
i56.14	1	1	1	0	0	0	0	0
i56.15	1	1	2	0	0	0	0	0
i56.16	1	1	1	1	0	0	0	0
i56.17	1	1	1	1	0	0	0	0

With respect to the monotonicity assumption, **Table 2.9** shows the results of the analysis, taking the three dimensions referred to above into account. **Table 2.9** also shows the homogeneity indices of each item in their dimensions and the homogeneity indices of each dimension. Significant violations (#zsig) of the monotonicity assumption are not observed for any Subscale III items in dimension 1. That is, all items of this dimension appear to discriminate well between respondents with high levels in the construct and those with lower levels. By contrast, however, violations of the monotonicity assumption were observed in dimension 2.

Table 2.9 – MSA – Monotonicity – Subscale III.I

Dimension 1 (H = 0.37, SE = 0.007)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i56.2	0.32	112	0	0	0	0	0	0	0	0
i56.3	0.32	112	0	0	0	0	0	0	0	0
i56.4	0.38	112	0	0	0	0	0	0	0	0
i56.5	0.4	112	0	0	0	0	0	0	0	0
i56.6	0.41	112	0	0	0	0	0	0	0	0
i56.7	0.42	112	0	0	0	0	0	0	0	0
i56.8	0.38	112	0	0	0	0	0	0	0	0
i56.10	0.36	112	0	0	0	0	0	0	0	0
i56.11	0.33	112	0	0	0	0	0	0	0	0
i56.12	0.38	112	0	0	0	0	0	0	0	0
i56.13	0.42	106	0	0	0	0	0	0	0	0
i56.16	0.31	112	0	0	0	0	0	0	0	0
i56.17	0.31	112	0	0	0	0	0	0	0	0

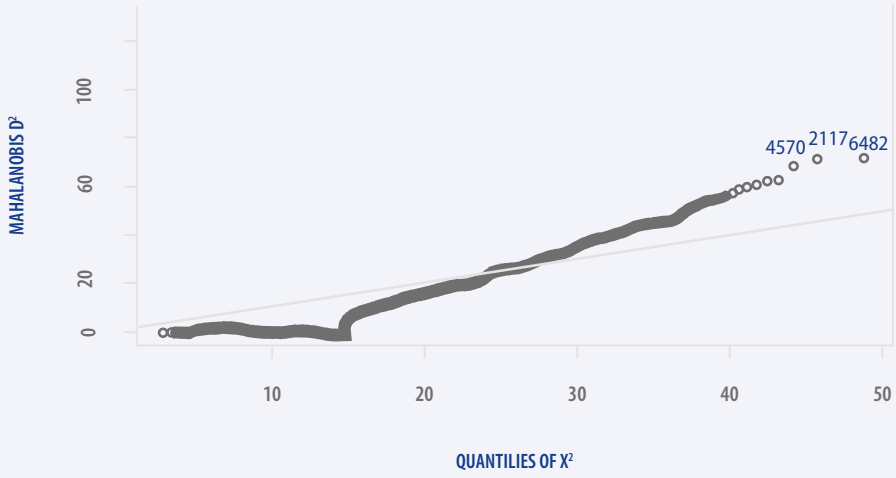
Dimension 2 (H = 0.16, SE = 0.009)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i56.1	0.14	211	0	0	0	0	0	0	0	0
i56.9	0.14	264	24	0.09	0.09	1.23	0.0047	3	12	101
i56.14	0.17	220	5	0.02	0.09	0.25	0.0011	1.98	1	46
i56.15	0.19	220	1	0	0.04	0.04	0.0002	1.29	0	19

d Multivariate outliers

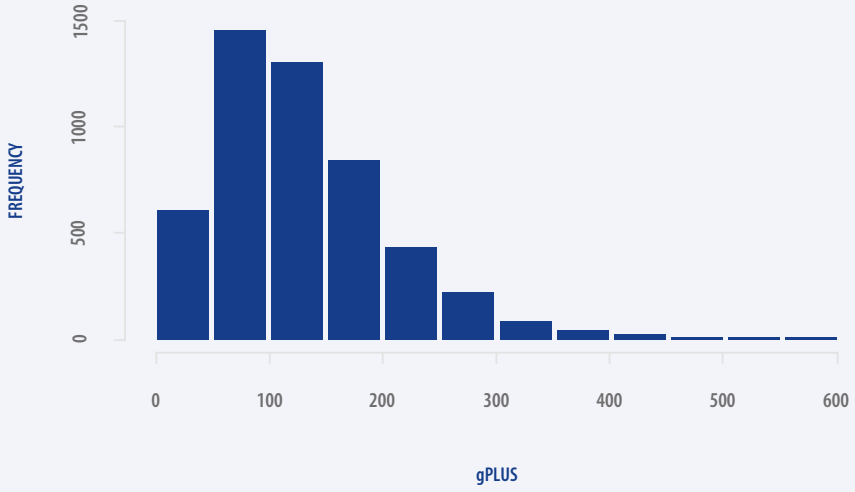
The multivariate outliers were then analysed by means of Mahalanobis D^2 distances. The results of the analysis are shown in **Figure 2.7**. As can be seen, the D^2 distances are shown on the y-axis while the chi-squared quantiles are shown on the x-axis. Of the D^2 distance values, 102 were significant at confidence level $\alpha = 0.001$ (Hair et al. 2019). The maximum D^2 value was 72.02.

Figure 2.7 – Q–Q Plot of Mahalanobis D^2 vs quantiles of χ^2 – Subscale III.I



The number of Guttman errors for each observation was also calculated to identify atypical response patterns. The error average was 103.03 (SD = 61.60); according to the criterion proposed by Zijlstra et al. (2007) and by Hubert and Vandervieren (2008) for asymmetric distributions, the critical value was 246.5, which was exceeded by 149 observations.

Figure 2.8 – Guttman error distribution – Subscale III.I



e Evidence of validity

Kaiser Meyer Olkin tests (Kaiser 1970) and Bartlett's sphericity test were used to test data adequacy for factor analysis. The result of the KMO test indicates that the data quality is adequate for factor analysis (KMO = 0.91). All items of this subscale obtained MSA values higher than or close to 0.7 (i56.1 = 0.92; i56.2 = 0.93; i56.3 = 0.92; i56.4 = 0.94; i56.5 = 0.89; i56.6 = 0.89; i56.7 = 0.88; i56.8 = 0.92; i56.9 = 0.68; i56.10 = 0.91; i56.11 = 0.91; i56.12 = 0.91; i56.13 = 0.91; i56.14 = 0.90; i56.15 = 0.86; i56.16 = 0.92; i56.17 = 0.92). The result of Bartlett's sphericity test was also significant (χ^2 (136) = 533.07; p < 0.001), the results obtained indicating that the data are adequate for submission to factor analysis.

The internal structure is shown in **Figure 2.9**. **Figure 2.10** shows the sedimentation graph with the result of exploratory factor analysis, which also supports the presence of two factors. The presence of one factor, however, is supported by six out of 27 methods (22.22%) (acceleration factor, Scree (R2), EGA (glasso), EGA (TMFG), VSS complexity 1, Velicer's MAP). It would be advisable to carry out a confirmatory factor analysis to compare the adjustment of the solutions for one and two factors proposed for the different methods, or bifactor or hierarchical models instead.

Figure 2.9 – Measurement module – Subscale III.I

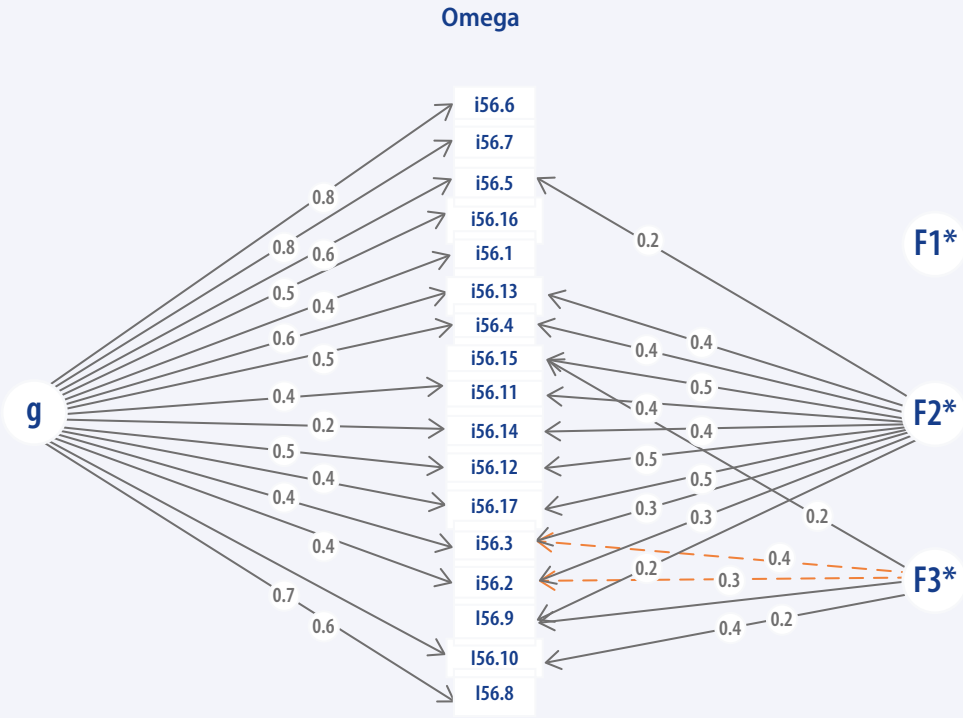
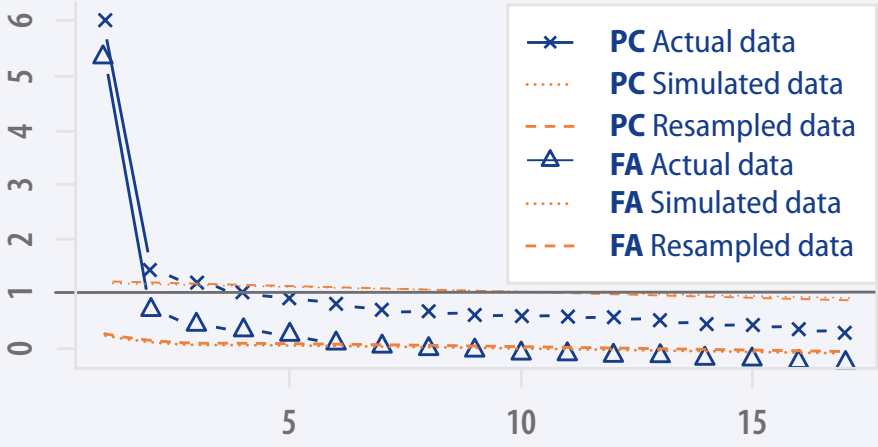


Figure 2.10 – Sedimentation graph – Subscale III.I



Subscale III.II: teacher views on history textbooks

a Descriptive analysis

The results of the descriptive analysis of the 12 items relating to the TES respondents' views on the history textbooks available in their countries (Subscale III, items 59.1 to 59.12) are shown in **Table 2.10**.

Table 2.10 – Level of agreement – Subscale III.II (items 59.1 to 59.12)

Item label	1	2	3	4	5	% 1 to 2	% 4 to 5	No. missing
i59.1	275	709	1 441	1 435	833	20.97	48.33	3 174
i59.2	1 039	923	1 345	977	409	41.81	29.53	3 174
i59.3	526	690	1 208	1 155	1 114	25.91	48.35	3 174
i59.4	653	769	1 332	1 175	764	30.3	41.32	3 174
i59.5	406	1 023	1 531	1 062	671	30.45	36.93	3 174
i59.6	1 076	1 154	1 270	699	494	47.52	25.42	3 174
i59.7	1 422	1 484	1 060	440	287	61.92	15.49	3 174
i59.8	617	1 125	1 210	908	833	37.12	37.1	3 174
i59.9	1 540	1 098	952	576	527	56.21	23.5	3 174
i59.10	1 788	999	872	462	572	59.39	22.03	3 174
i59.11	1 172	1 046	1 369	729	377	47.26	23.57	3 174
i59.12	435	887	1 509	1 100	762	28.17	39.68	3 174

Item	n	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
i59.1	4693	3.39	1.12	3	1	5	4	-0.3	-0.61	0.02
i59.2	4693	2.74	1.25	3	1	5	4	0.08	-1.03	0.02
i59.3	4693	3.35	1.29	3	1	5	4	-0.32	-0.95	0.02
i59.4	4693	3.13	1.27	3	1	5	4	-0.18	-0.95	0.02
i59.5	4693	3.12	1.16	3	1	5	4	-0.02	-0.79	0.02
i59.6	4693	2.66	1.27	3	1	5	4	0.3	-0.91	0.02
i59.7	4693	2.29	1.17	3	1	5	4	0.68	-0.34	0.02
i59.8	4693	3.05	1.29	3	1	5	4	0.04	-1.08	0.02
i59.9	4693	2.46	1.35	3	1	5	4	0.52	-0.93	0.02
i59.10	4693	2.37	1.39	3	1	5	4	0.64	-0.86	0.02
i59.11	4693	2.59	1.24	3	1	5	4	0.27	-0.9	0.02
i59.12	4693	3.18	1.19	3	1	5	4	-0.11	-0.81	0.02

b Reliability

With respect to the reliability of this subscale of 12 items, values higher than 0.8 were obtained through both Cronbach's ordinal alpha (α = 0.86) and McDonald's omega (ω = 0.9) (McDonald 2013; Revelle and Zinbarg 2009). Values higher than 0.8 are considered to be good (Kline 1999). **Table 2.11** presents the results of the reliability analysis and the item–total correlations of the scale.

It can be seen that all item–total correlations were greater than 0.3. Items i59.2, i59.4 and i59.11 were negatively correlated with total scale and were reversed.

Table 2.11 – Reliability analysis – Subscale III.II

Item	Alpha if an item is dropped	Item–total correlation
i59.1	0.84	0.634
i59.2	0.86	0.331
i59.3	0.86	0.430
i59.4	0.88	0.052
i59.5	0.84	0.657
i59.6	0.84	0.626
i59.7	0.84	0.614
i59.8	0.83	0.782
i59.9	0.84	0.703
i59.10	0.84	0.644
i59.11	0.86	0.328
i59.12	0.84	0.698

c Mokken scale analysis

With respect to item homogeneity, the general scalability coefficient obtained for the 12 items was H = 0.222 (SE = 0.005). The H scalability values of all items are shown in **Table 2.12**.

Table 2.12 – Homogeneity coefficients – Subscale III.II

Item	H	SE
i59.1	0.278	(0.007)
i59.2	-0.098	(0.010)
i59.3	0.220	(0.008)
i59.4	0.035	(0.010)
i59.5	0.307	(0.007)
i59.6	0.334	(0.007)
i59.7	0.346	(0.007)
i59.8	0.354	(0.006)
i59.9	0.351	(0.007)
i59.10	0.331	(0.007)
i59.11	-0.100	(0.010)
i59.12	0.300	(0.007)

These figures once again indicate the possible multidimensionality of the scale. The automated item selection procedure was then carried out at increasing homogeneity threshold levels to examine dimensionality. As has been shown, if all items appear as belonging to dimension 1, this indicates that the scale is unidimensional within that homogeneity threshold (indicated in the column headings, from 0.1 to 0.5). **Table 2.13** shows the results of the AISP: two dimensions are identified in the item set with a homogeneity threshold of $H \geq 0.3$. All items except i59.2 and i59.11 would remain in dimension 1. Only i59.4 does not appear to fit in either dimension. These results confirm the presence of a multidimensional structure in the scale that will have to be examined by means of exploratory factor analysis.

Table 2.13 – MSA – AISP for increasing H thresholds *(t)* – Subscale III.II

Item	*t* = 0.10	*t* = 0.15	*t* = 0.20	*t* = 0.30	*t* = 0.35	*t* = 0.40	*t* = 0.45	*t* = 0.50
i59.1	1	1	1	1	1	1	1	2
i59.2	2	2	2	2	0	0	0	0
i59.3	1	1	1	1	0	0	0	0
i59.4	2	2	0	0	0	0	0	0
i59.5	1	1	1	1	1	1	1	2
i59.6	1	1	1	1	1	1	1	1
i59.7	1	1	1	1	1	1	1	1
i59.8	1	1	1	1	1	1	1	1
i59.9	1	1	1	1	1	1	1	1
i59.10	1	1	1	1	1	1	1	1
i59.11	2	2	2	2	0	0	0	0
i59.12	1	1	1	1	1	1	1	2

With respect to the monotonicity assumption, **Table 2.14** shows the results of the analysis, taking the two dimensions referred to above into account. **Table 2.14** also presents the homogeneity indices of each item in their dimensions and the homogeneity indices of each dimension. Significant violations (#zsig) of the monotonicity assumption are not observed for any of the dimension 1 items of Subscale III.II, though a monotonicity violation was observed in item i59.2 of dimension 2. Thus all dimension 1 items appear to discriminate clearly between respondents with high levels in the construct and those with lower levels, though this is not the case of item i59.2 in dimension 2.

Table 2.14 – MSA – Monotonicity – Subscale III.II

Dimension 1 (H = 0.48, SE = 0.007)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i59.1	0.45	112	0	0	0	0	0	0	0	0
i59.3	0.33	112	0	0	0	0	0	0	0	0
i59.5	0.48	112	0	0	0	0	0	0	0	0
i59.6	0.49	112	0	0	0	0	0	0	0	0
i59.7	0.49	112	0	0	0	0	0	0	0	0
i59.8	0.56	112	0	0	0	0	0	0	0	0
i59.9	0.53	112	0	0	0	0	0	0	0	0
i59.10	0.49	112	0	0	0	0	0	0	0	0
i59.12	0.49	112	0	0	0	0	0	0	0	0

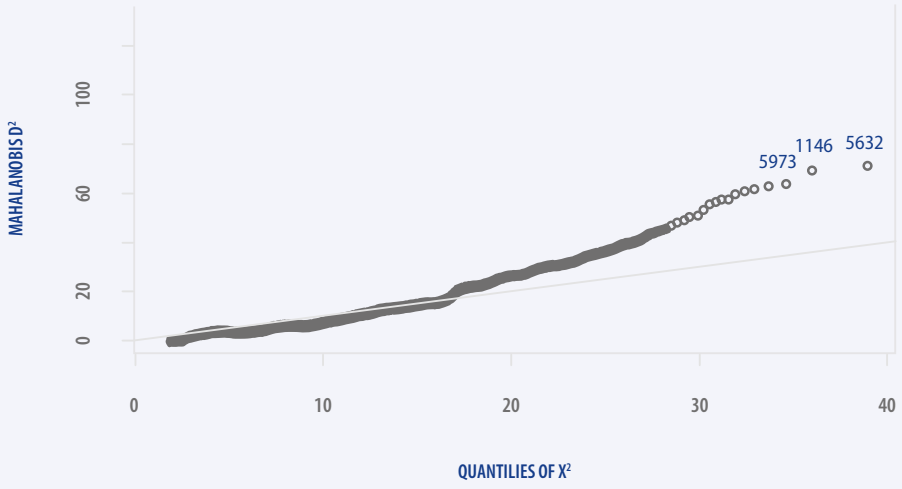
Dimension 2 (H = 0.33, SE = 0.015)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i59.2	0.33	24	1	0.04	0.04	0.04	0.0015	3.95	1	44
i59.11	0.33	24	0	0.00	0.00	0.00	0.0000	0.00	0	0

d Multivariate outliers

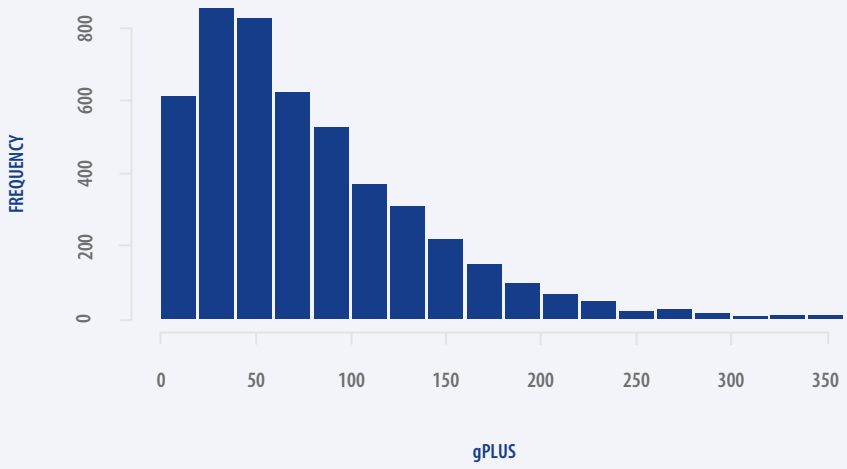
The multivariate outliers were then analysed by means of the Mahalanobis D^2 distances. The results of the analysis are shown in **Figure 2.11**. As can be seen, the D^2 distances are shown on the y-axis while the chi-squared quantiles are shown on the x-axis. Of the D^2 distance values, 101 were significant at confidence level $\alpha = 0.001$ (Hair et al. 2019). The maximum D^2 value was 71.09.

Figure 2.11 – Q–Q Plot of Mahalanobis D^2 vs quantiles of χ^2 – Subscale III.II



The number of Guttman errors for each observation was also calculated to identify atypical response patterns. The error average was 74.51 (SD = 53.4); according to the criterion proposed by Zijlstra et al. (2007) and by Hubert and Vandervieren (2008) for asymmetric distributions, the critical value was 210.5, which was exceeded by 110 observations.

Figure 2.12 – Guttman error distribution – Subscale III.II



e Evidence of validity

To verify data adequacy for factor analysis, the Kaiser Meyer Olkin test (KMO) (Kaiser 1970) and Bartlett spherical tests were used. The result of the KMO test indicates that the data are adequate for factor analysis (KMO = 0.88). All items of this subscale obtained MSA values higher than 0.7 (i59.1 = 0.88; i59.2 = 0.82; i59.3 = 0.93; i59.5 = 0.93; i59.6 = 0.87; i59.7 = 0.88; i59.8 = 0.93; i59.9 = 0.86; i59.10 = 0.86; i59.11 = 0.83; i59.12 = 0.90), except item i59.4 (MSA = 0.52). The result of Bartlett's sphericity test was also significant (χ^2 (66) = 26 831.58; p < 0.001). The results obtained indicated that the data are adequate for factor analysis.

The internal structure is shown in **Figure 2.13**.

Figure 2.14 presents the sedimentation graph with the result of exploratory factor analysis, which also supports the presence of two factors. However, the presence of three factors is supported by seven methods out of 27 (25.93%) (CNG, optimal co-ordinates, parallel analysis, Kaiser criterion, EGA (glasso), EGA (TMFG), VSS complexity 2). It would be advisable to carry out a confirmatory factor analysis and to revise item i59.4.

Figure 2.13 – Measurement module – Subscale III.II

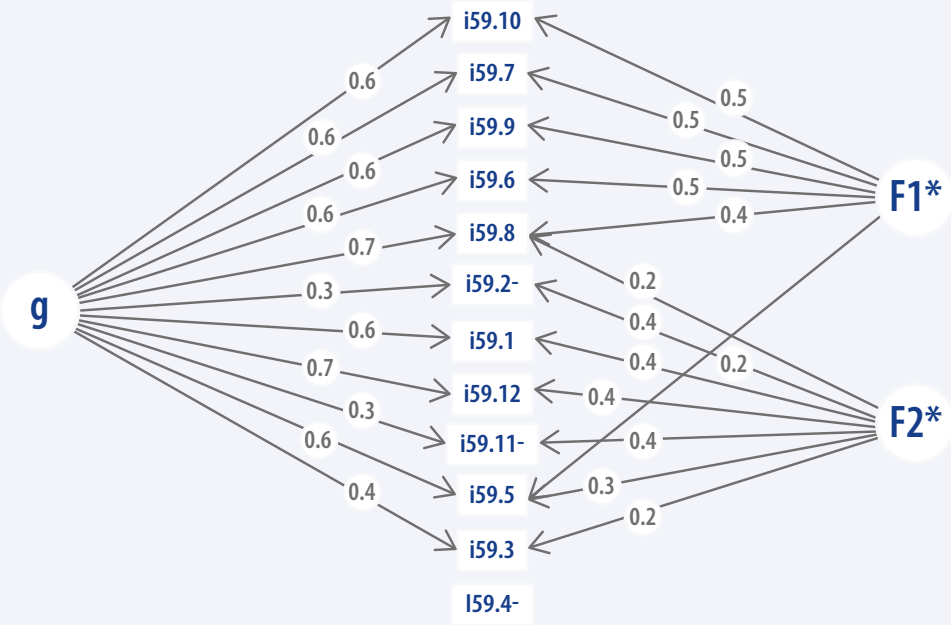
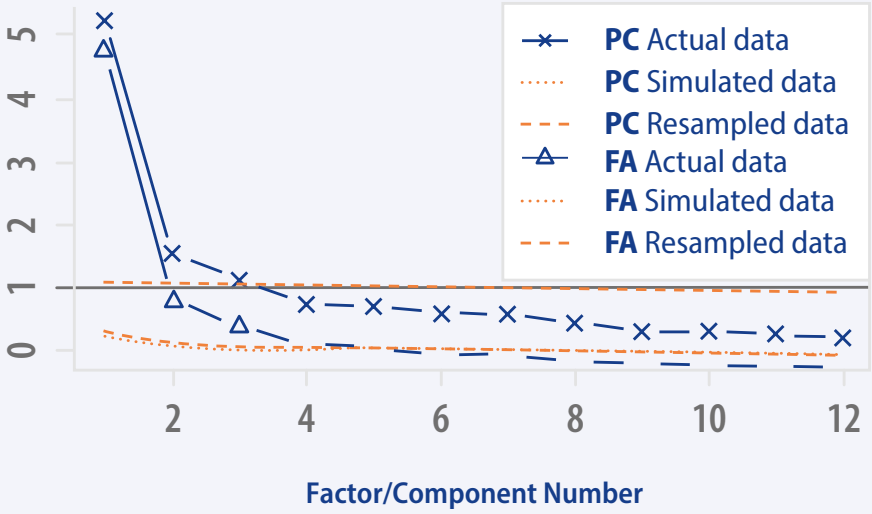


Figure 2.14 – Sedimentation graph – Subscale III.II



SUBSCALE IV – History teaching and learning in practice

Subscale IV.I: barriers to quality history teaching

a Descriptive analysis

The results of the descriptive analysis of the eight items relating to the frequency of use of the different methods for teaching and learning history (Subscale IV, items 62.1 to 62.8) are shown in **Table 2.15**.

Table 2.15 – Frequency of use of methods for teaching and learning history – Subscale IV.I: barriers to quality history teaching

Item	Item label	1	2	3	4	5	% 1 to 2	% 4 to 5	Missing
Delivering lectures/ presentations	i62.1	140	343	964	1 556	1 534	10.65	68.11	3 330
Debating on controversial issues	i62.2	124	550	1 416	1 520	927	14.86	53.93	3 330

Item	Item label	1	2	3	4	5	% 1 to 2	& 4 to 5	Missing
Questioning how history is represented in the public space (movies, street names, monuments, games, graphic novels, etc.)	i62.3	189	754	1 570	1 368	656	20.78	44.61	3 330
Reflecting on how history is written and used	i62.4	144	647	1 642	1 406	698	17.43	46.37	3 330
Project-based learning	i62.5	282	905	1 549	1 263	538	26.16	39.7	3 330
Place-based learning (outside the classroom, such as visits to museums, historical sites and archives)	i62.6	426	1 434	1 484	887	306	41	26.29	3 330
Working with periodisations and timelines	i62.7	130	574	1 367	1 393	1 073	15.52	54.35	3 330
Using contrasting historical sources and multiple narratives about past events	i62.8	177	671	1 477	1 410	802	18.69	48.75	3 330

Item	n	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
Delivering lectures/ presentations	4 537	3.88	1.06	4	1	5	4	-0.77	-0.01	0.02
Debating on controversial issues	4 537	3.57	1.03	4	1	5	4	-0.33	-0.49	0.02
Questioning how history is represented in the public space (movies, street names, monuments, games, graphic novels, etc.)	4 537	3.34	1.05	3	1	5	4	-0.18	-0.18	0.02

Item	n	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
Reflecting on how history is written and used	4537	3.41	1.01	3	1	5	4	-0.18	-0.47	0.02
Project-based learning	4537	3.19	1.08	3	1	5	4	-0.12	-0.62	0.02
Place-based learning (outside the classroom, such as visits to museums, historical sites and archives)	4537	2.83	1.06	3	1	5	4	0.22	-0.58	0.02
Working with periodisations and timelines	4537	3.6	1.07	4	1	5	4	-0.33	-0.62	0.02
Using contrasting historical sources and multiple narratives about past events	4537	3.44	1.06	3	1	5	4	-0.25	-0.57	0.02

b Reliability

With respect to the reliability of this range of eight items, values higher than 0.8 were obtained by means of both Cronbach's ordinal alpha (α = 0.82) and McDonald's omega (ω = 0.87) (McDonald 2013; Revelle and Zinbarg 2009). Values higher than 0.8 are considered to be good (Kline 1999). **Table 2.16** presents the results of the reliability analysis and the item–total correlations of the scale.

It can be seen that all item–total correlations were greater than 0.3.

Table 2.16 – Reliability analysis – Subscale IV.I

Item	Alpha if an item is dropped	Item–total correlation
i62.1	0.83	0.32
i62.2	0.79	0.63
i62.3	0.79	0.65
i62.4	0.79	0.64
i62.5	0.80	0.55
i62.6	0.81	0.49
i62.7	0.80	0.53
i62.8	0.80	0.55

c Mokken scale analysis

With respect to item homogeneity, the general scalability coefficient obtained for the eight items was H = 0.357 (SE = 0.007). The H scalability values of all items is shown in **Table 2.17**.

Table 2.17 – Homogeneity coefficients – Subscale IV.I

Item	H	SE
i62.1	0.217	(0.011)
i62.2	0.404	(0.009)
i62.3	0.416	(0.009)
i62.4	0.413	(0.009)
i62.5	0.365	(0.010)
i62.6	0.337	(0.010)
i62.7	0.343	(0.010)
i62.8	0.357	(0.010)

All values except for item 62.1 exceed threshold H = 0.30. Multidimensionality indicators were therefore not identified and the items are scalable to H ≥ 0.30, which indicates average accuracy (Stochl et al. 2012). However, it would be appropriate to review the inclusion of item 62.1 in this scale.

The automated item selection procedure was then carried out at increasingly homogeneous threshold levels to examine dimensionality. As has been shown, if all items appear as belonging to dimension 1, this indicates that the scale is unidimensional in that homogeneity threshold (indicated in the column headings, from 0.1 to 0.5). **Table 2.18** shows the results of the AISP.

As has been shown, all items except 62.1 form part of a unique dimension with a homogeneity threshold of H ≥ 0.3. These results appear to confirm the presence of a unidimensional structure in the scale that will have to be tested by means of exploratory factor analysis.

Table 2.18 – MSA – AISP for increasing H thresholds *(t)* – Subscale IV.I

Item	*t* = 0.10	*t* = 0.15	*t* = 0.20	*t* = 0.30	*t* = 0.35	*t* = 0.40	*t* = 0.45	*t* = 0.50
i62.1	1	1	1	0	0	0	0	0
i62.2	1	1	1	1	1	1	1	1
i62.3	1	1	1	1	1	1	1	1
i62.4	1	1	1	1	1	1	1	1
i62.5	1	1	1	1	1	2	2	0
i62.6	1	1	1	1	1	2	2	0
i62.7	1	1	1	1	1	0	3	0
i62.8	1	1	1	1	1	1	3	0

With respect to the monotonicity assumption, **Table 2.19** shows the results of the analysis. It also presents the homogeneity indices of each item. No significant violations (#zsig) of the monotonicity assumption are observed for any Subscale IV items. That is, all items appear to discriminate well between respondents with high levels in the construct and those with lower levels.

Table 2.19 – MSA – Monotonicity – Subscale IV.I

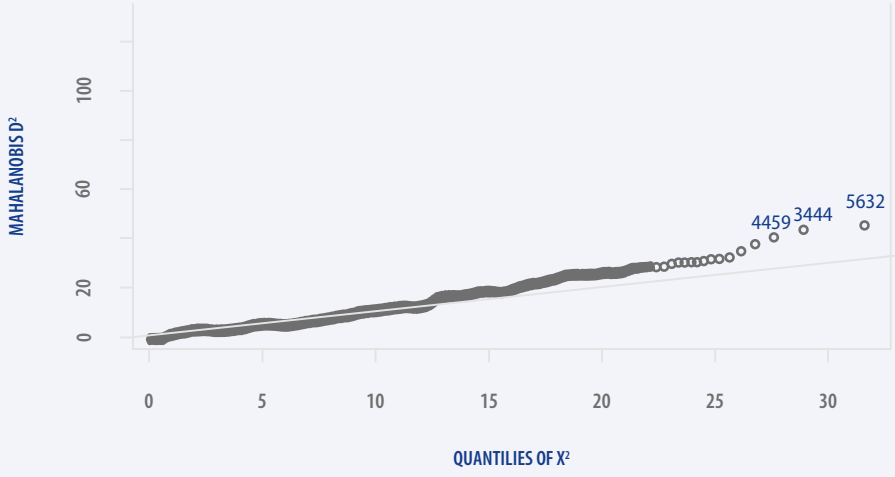
Dimension 1 (H = 0.40, SE = 0.008)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i62.2	0.41	84	0	0	0	0	0	0	0	0
i62.3	0.44	84	0	0	0	0	0	0	0	0
i62.4	0.44	84	0	0	0	0	0	0	0	0
i62.5	0.40	84	0	0	0	0	0	0	0	0
i62.6	0.37	84	0	0	0	0	0	0	0	0

d Multivariate outliers

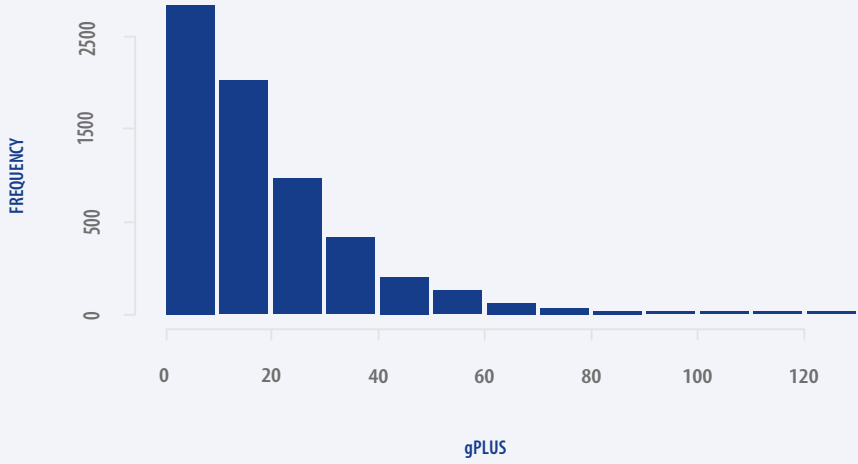
The multivariate outliers were then analysed by means of the Mahalanobis D^2 distances. The results of the analysis are shown in **Figure 2.15**. As can be seen, the D^2 distances are shown on the y-axis while the chi-squared quantiles are shown on the x-axis. Of the D^2 distance values, 36 were significant at confidence level $\alpha = 0.001$ (Hair et al. 2019). The maximum D^2 value was 45.07.

Figure 2.15 – Q–Q Plot of Mahalanobis D^2 vs quantiles of χ^2 – Subscale IV.I



The number of Guttman errors for each observation was also calculated to identify atypical response patterns. The error average was 18.30 (SD = 15.56); according to the criterion proposed by Zijlstra et al. (2007) and Hubert and Vandervieren (2008) for asymmetric distributions, the critical value was 52, which was exceeded by 105 observations.

Figure 2.16 – Guttman error distribution – Subscale IV.I



e Evidence of validity

To verify data adequacy for factor analysis, the Kaiser Meyer Olkin (Kaiser 1970) and Bartlett spherical tests were used. The result of the KMO test indicates that the data are adequate for factor analysis (KMO = 0.85). All items of this subscale obtained MSA values in excess of 0.7 (i62.1 = 0.79; i62.2 = 0.85; i62.3 = 0.85; i62.4 = 0.86; i62.5 = 0.87; i62.6 = 0.86; i62.7 = 0.85; i62.8 = 0.85). The result of Bartlett's sphericity test was also significant (χ^2 (28) = 11 021.08; $p < 0.001$). The results obtained indicated that the data are adequate for submission to factor analysis.

The internal structure is shown in **Figure 2.17**.

Figure 2.18 presents the sedimentation graph with the result of exploratory factor analysis, which also suggests the presence of a single factor. Similarly, the presence of a single factor is supported by 12 methods out of 27 (44.44%) (t, p, optimal co-ordinates, acceleration factor, parallel analysis, Kaiser criterion, Scree (SE), Scree (R2), EGA (glasso), EGA (TMFG), VSS complexity 1, Velicer's MAP).

Figure 2.17 – Measurement module – Subscale IV.I

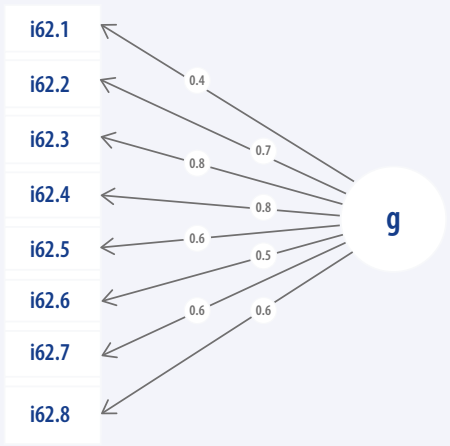
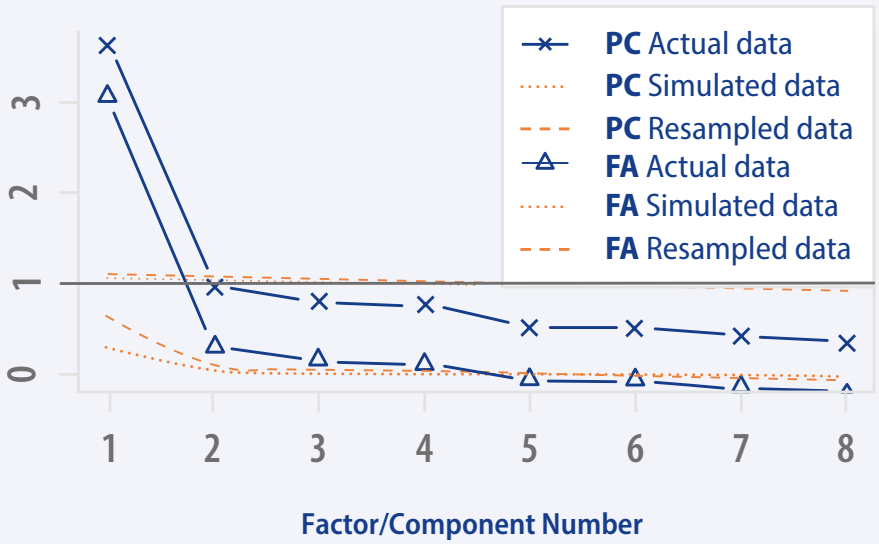


Figure 2.18 – Sedimentation graph – Subscale IV.I



Subscale IV.II: content focus (importance)

a Descriptive analysis

The results of the descriptive analysis of the seven items relating to the importance of different themes in history teaching (Subscale IV.II, items 65.1 to 65.7) are shown in **Table 2.20**.

Table 2.20 – Frequency of use of different resources – Subscale IV.II (importance)

Item	Item label	1	2	3	4	5	% 1 to 2	% 4 to 5	Missing
Art history	i65.1	124	585	1 208	1 192	1 170	16.57	55.2	3 588
Social and economic history	i65.2	43	243	840	1 379	1 774	6.68	73.69	3 588
Political and military history	i65.3	64	274	837	1 291	1 813	7.9	72.54	3 588
Gender history	i65.4	523	901	1 289	908	658	33.28	36.6	3 588
History of minorities and cultures	i65.5	125	599	1 300	1 288	967	16.92	52.7	3 588
Migration history	i65.6	66	463	1 295	1 390	1 065	12.36	57.37	3 588
Environmental history	i65.7	182	628	1 217	1 163	1 089	18.93	52.63	3 588

Item	n	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
Art history	4 279	3.63	1.11	4	1	5	4	-0.35	-0.77	0.02
Social and economic history	4 279	4.07	0.96	4	1	5	4	-0.81	-0.05	0.01
Political and military history	4 279	4.06	1	4	1	5	4	-0.85	-0.02	0.02
Gender history	4 279	3.06	1.23	3	1	5	4	-0.02	-0.93	0.02
History of minorities and cultures	4 279	3.55	1.07	4	1	5	4	-0.29	-0.68	0.02
Migration history	4 279	3.68	1.01	4	1	5	4	-0.32	-0.63	0.02
Environmental history	4 279	3.55	1.14	4	1	5	4	-0.33	-0.77	0.02

b Reliability

With respect to the reliability of this set of seven items, values higher than 0.8 were obtained through both Cronbach's ordinal alpha (α = 0.83) and McDonald's omega (ω = 0.89) (McDonald 2013; Revelle and Zinbarg 2009). Values higher than 0.8 are considered to be good (Kline 1999). **Table 2.21** presents the results of the reliability analysis and the item–total correlations of the scale.

It can be seen that all item–total correlations were greater than 0.3.

Table 2.21 – Reliability analysis – Subscale IV.II (importance)

Item	Alpha if an item is dropped	Item–total correlation
i65.1	0.81	0.55
i65.2	0.81	0.59
i65.3	0.84	0.34
i65.4	0.81	0.58
i65.5	0.79	0.71
i65.6	0.79	0.69
i65.7	0.81	0.59

Table 2.22 – Homogeneity coefficients – Subscale IV.II (importance)

Item	H	SE
i65.1	0.365	(0.010)
i65.2	0.391	(0.010)
i65.3	0.227	(0.012)
i65.4	0.407	(0.010)
i65.5	0.476	(0.009)
i65.6	0.467	(0.009)
i65.7	0.402	(0.010)

c Mokken scale analysis

With respect to item homogeneity, the general scalability coefficient obtained for the eight items was H = 0.357 (SE = 0.007). The H scalability values of all items are shown in **Table 2.22**.

All values except for item 65.3 exceed the threshold of H = 0.30. Multidimensionality indicators were therefore not identified and the items are scalable to H ≥ 0.30, which indicates average accuracy (Stochl et al. 2012). However, it would be appropriate to review the inclusion of item 65.3 in this scale.

The automated item selection procedure was then carried out at increasing homogeneity threshold levels to examine dimensionality. As has been shown, if all items appear to belong to dimension 1, this indicates that the scale is unidimensional within that homogeneity threshold (indicated in the column headings, from 0.1 to 0.5). **Table 2.23** shows the results of the AISP. As can be seen, all items except for 65.3 appear to form part of a single dimension with a homogeneity threshold of H ≥ 0.3. These results appear to confirm the presence of a unidimensional structure in the scale that will have to be proven by exploratory factor analysis.

Table 2.23 – MSA – AISP for increasing H thresholds *(t)* – Subscale IV.II (importance)

Item	*t* = 0.10	*t* = 0.15	*t* = 0.20	*t* = 0.30	*t* = 0.35	*t* = 0.40	*t* = 0.45	*t* = 0.50
i65.1	1	1	1	1	1	2	2	2
i65.2	1	1	1	1	1	2	2	2
i65.3	1	1	1	0	0	0	0	0
i65.4	1	1	1	1	1	1	1	1
i65.5	1	1	1	1	1	1	1	1
i65.6	1	1	1	1	1	1	1	1
i65.7	1	1	1	1	1	1	1	1
i65.8	1	1	1	1	1	1	1	1

With respect to the monotonicity assumption, **Table 2.24** shows the results of the analysis. The homogeneity indices of each item are also shown in **Table 2.24**. Significant violations (#zsig) of the monotonicity assumption for any items of Subscale IV.II, content focus (importance), were not observed. That is, all items appear to discriminate well between respondents with high levels in the construct and those with lower levels.

Table 2.24 – MSA – Monotonicity – Subscale IV.II (importance)

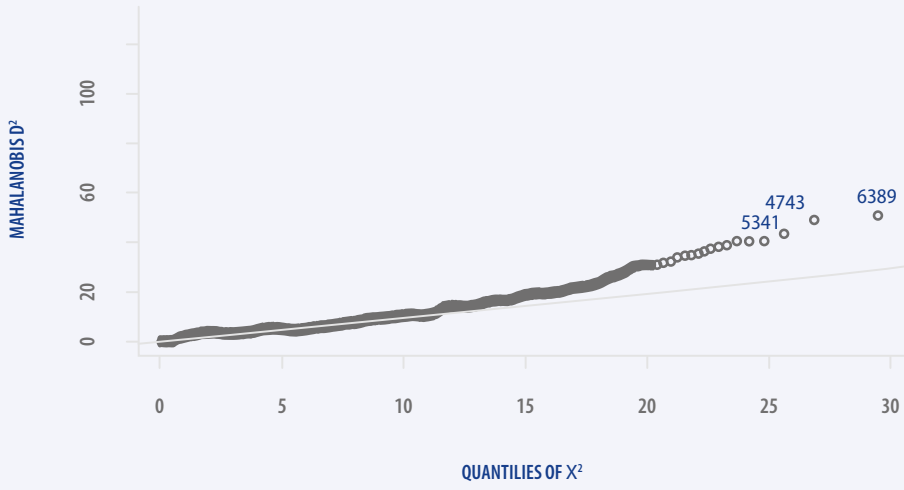
Dimension 1 (H = 0.45, SE = 0.008)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i65.1	0.39	60	0	0	0	0	0	0	0	0
i65.2	0.39	84	0	0	0	0	0	0	0	0
i65.4	0.45	84	0	0	0	0	0	0	0	0
i65.5	0.52	60	0	0	0	0	0	0	0	0
i65.6	0.50	50	0	0	0	0	0	0	0	0
i65.7	0.45	84	0	0	0	0	0	0	0	0

d Multivariate outliers

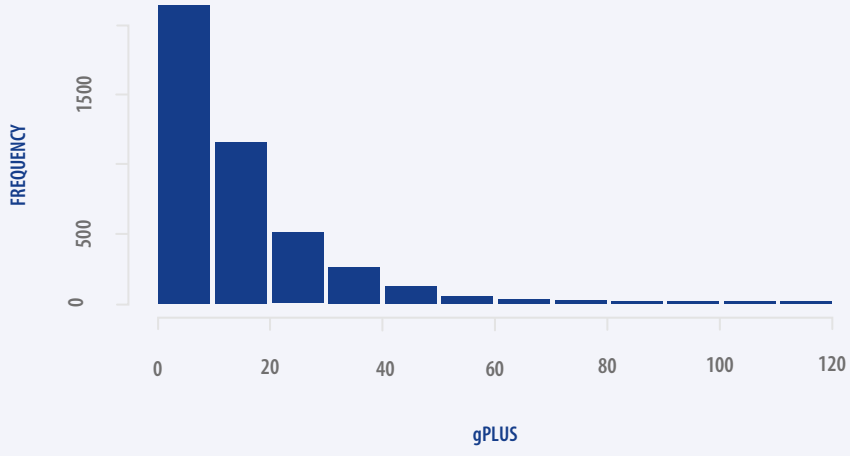
The multivariate outliers were then analysed by means of the Mahalanobis D^2 distances. The results of the analysis are shown in **Figure 2.19**. As can be seen, the D^2 distances are shown on the y-axis while the chi-squared quantiles are shown on the x-axis. Of the D^2 distance values, 49 were significant at confidence level $\alpha = 0.001$ (Hair et al. 2019). The maximum D^2 value was 51.09.

Figure 2.19 – Q–Q Plot of Mahalanobis D^2 vs quantiles of χ^2 – Subscale IV.II (importance)



The number of Guttman errors for each observation was also calculated to identify atypical response patterns. The error average was 13.61 (SD = 13.01); according to the criterion proposed by Zijlstra et al. (2007) and Hubert and Vandervieren (2008) for asymmetric distributions, the critical figure was 41.5, which was exceeded by 179 observations.

Figure 2.20 – Guttman error distribution – Subscale IV.II (importance)



e Evidence of validity

The Kaiser Meyer Olkin (Kaiser 1970) tests and Bartlett's sphericity tests were used to verify data adequacy for factor analysis. The result of the KMO test indicates that the data are adequate for factor analysis (KMO = 0.82). All items of this subscale obtained MSA values higher than 0.7 (i65.1 = 0.82; i65.2 = 0.77; i65.3 = 0.73; i65.4 = 0.85; i65.5 = 0.82; i65.6 = 0.82; i65.7 = 0.85). The result of Bartlett's sphericity test was also significant (χ^2 (21) = 12517.0; p< 0.001). The results obtained suggested that the data are adequate for submission to factor analysis.

The internal structure is shown in **Figure 2.21**.

Figure 2.22 shows the sedimentation graph with the result of the exploratory factor analysis, which also suggests the presence of a single factor. The presence of two factors, however, is supported by seven out of 24 methods (29.17%) (optimal co-ordinates, parallel analysis, Kaiser criterion, EGA (glasso), EGA (TMFG), VSS complexity 2, CRMS), and it would therefore be appropriate to carry out an in-depth analysis of the internal structure of the scale by means of a confirmatory factor analysis.

Figure 2.21 – Measurement module – Subscale IV.II (importance)

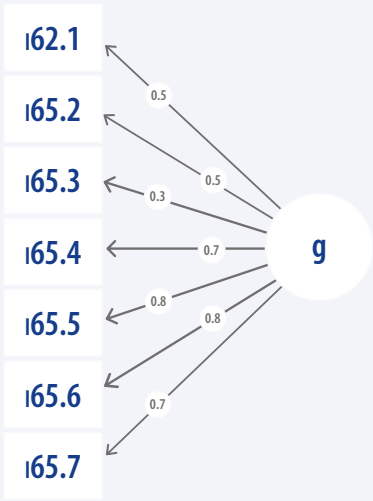
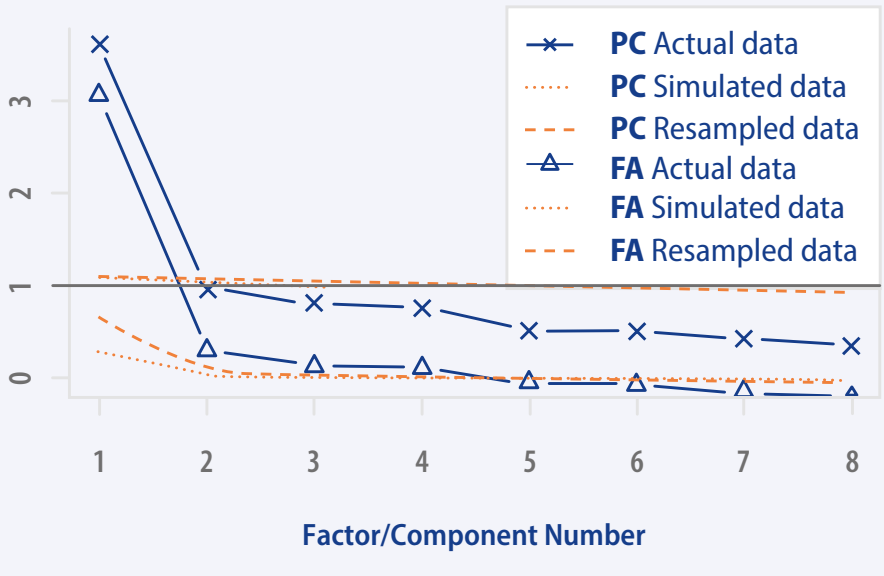


Figure 2.22 – Sedimentation graph – Subscale IV.II (importance)



Subscale IV.II: content focus (frequency)

a Descriptive analysis

The results of the descriptive analysis of the seven items relating to the frequency of teaching of the different themes in history (Subscale IV.II – items 66.1 to 66.7) are shown in **Table 2.25**.

Table 2.25 – Frequency of use of different resources – Subscale IV.II (frequency)

Item	Item label	1	2	3	4	5	% 1 to 2	% 4 to 5	Missing
Art history	i66.1	340	1 031	1 313	940	623	32.28	36.8	3 620
Social and economic history	i66.2	144	466	1 033	1 342	1 262	14.36	61.31	3 620
Political and military history	i66.3	151	318	664	1 192	1 922	11.04	73.32	3 620
Gender history	i66.4	961	1 324	1 109	539	314	53.8	20.08	3 620
History of minorities and cultures	i66.5	400	1 186	1 267	832	562	37.34	32.82	3 620
Migration history	i66.6	437	1 199	1 420	774	417	38.52	28.04	3 620
Environmental history	i66.7	824	1 341	1 086	636	360	50.98	23.45	3 620

Item	n	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
Art history	4 247	3.11	1.17	3	1	5	4	0.03	-0.85	0.02
Social and economic history	4 247	3.73	1.1	4	1	5	4	-0.56	-0.47	0.02
Political and military history	4 247	4.04	1.11	4	1	5	4	-1.04	0.25	0.02
Gender history	4 247	2.51	1.18	2	1	5	4	0.46	-0.61	0.02
History of minorities and cultures	4 247	2.99	1.18	3	1	5	4	0.15	-0.85	0.02
Migration history	4 247	2.89	1.12	3	1	5	4	0.2	-0.65	0.02
Environmental history	4 247	2.62	1.2	2	1	5	4	0.39	-0.73	0.02

b Reliability

With respect to the reliability of this group of seven items, values higher than 0.8 were obtained by means of both Cronbach's ordinal alpha (α = 0.84) and McDonald's omega (ω 0.91) (McDonald 2013; Revelle and Zinbarg 2009). Values higher than 0.8 are considered to be good (Kline 1999). **Table 2.26** presents the results of the reliability analysis and the item–total correlations of each scale.

It can be seen that all item–total correlations were greater than 0.3.

Table 2.26 – Reliability analysis – Subscale IV.II (frequency)

Item	Alpha if an item is dropped	Item–total correlation
i66.1	0.81	0.61
i66.2	0.81	0.63
i66.3	0.85	0.39
i66.4	0.82	0.58
i66.5	0.81	0.62
i66.6	0.80	0.67
i66.7	0.81	0.62

c Mokken scale analysis

With respect to item homogeneity, the general scalability coefficient obtained for the seven items was H = 0.421 (SE = 0.008). The H scalability values of all items are shown in **Table 2.27**.

All values except for item 66.3 exceed the threshold H = 0.30. Multidimensionality indicators were therefore not identified and the items are scalable to H ≥ 0.30, which indicates average accuracy (Stochl et al. 2012). It would be useful, however, to review the inclusion of item 66.3 in this scale.

Table 2.27 – Homogeneity coefficients – Subscale IV.II (frequency)

Item	H	SE
i66.1	0.428	(0.010)
i66.2	0.437	(0.011)
i66.3	0.292	(0.011)
i66.4	0.423	(0.010)
i66.5	0.440	(0.012)
i66.6	0.469	(0.010)
i66.7	0.445	(0.009)

The automated item selection procedure was then carried out at increasingly homogeneous threshold levels to examine dimensionality. As has been shown, if all items appear to belong to dimension 1, this indicates that the scale is unidimensional within that homogeneity threshold (indicated in the column headings, from 0.1 to 0.5). **Table 2.28** shows the results of the AISP. All items except for 66.3 appear to form part of a unique dimension with a homogeneity threshold of H ≥ 0.3. These results appear to confirm the presence of a unidimensional structure in the scale that will have to be proven by means of exploratory factor analysis.

Table 2.28 – MSA – AISP for increasing H thresholds *(t)* – Subscale IV.II (frequency)

Item	*t* = 0.10	*t* = 0.15	*t* = 0.20	*t* = 0.30	*t* = 0.35	*t* = 0.40	*t* = 0.45	*t* = 0.50
i66.1	1	1	1	1	1	1	0	0
i66.2	1	1	1	1	1	2	2	2
i66.3	1	1	1	0	0	2	2	2
i66.4	1	1	1	1	1	1	1	1
i66.5	1	1	1	1	1	1	1	1
i66.6	1	1	1	1	1	1	1	1
i66.7	1	1	1	1	1	1	1	1

With respect to the monotonicity assumption, **Table 2.29** shows the results of the analysis. This table also presents the homogeneity indices of each item. Significant violations (#zsig) of the monotonicity assumption are not observed for any Subscale IV.II items: content focus (frequency).

Table 2.29 – MSA – Monotonicity – Subscale IV.II (frequency)

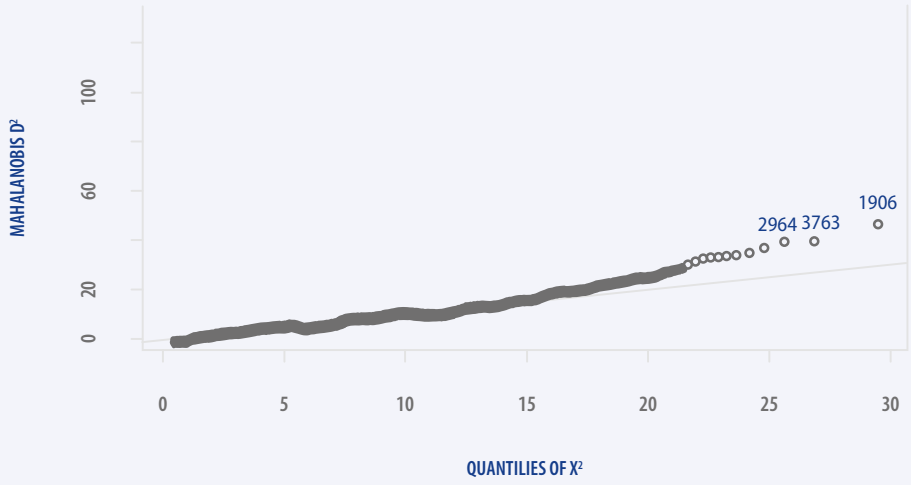
Dimension 1 (H = 0.40, SE = 0.009)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i66.1	0.45	84	0	0.00	0.00	0.00	0	0.00	0	0
i66.2	0.40	60	0	0.00	0.00	0.00	0	0.00	0	0
i66.4	0.47	84	0	0.00	0.00	0.00	0	0.00	0	0
i66.5	0.48	84	0	0.00	0.00	0.00	0	0.00	0	0
i66.6	0.49	84	0	0.00	0.00	0.00	0	0.00	0	0
i66.7	0.50	84	0	0.00	0.00	0.00	0	0.00	0	0

d Multivariate outliers

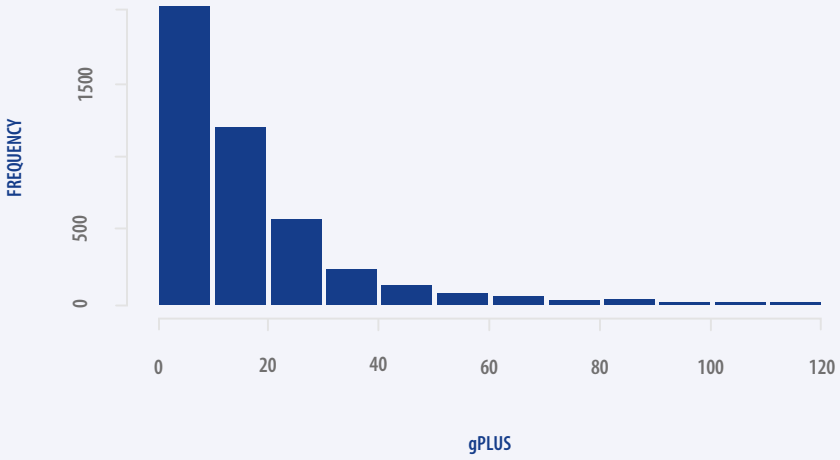
The multivariate outliers were then analysed by means of the Mahalanobis D^2 distances. The results of the analysis are shown in **Figure 2.23**. As can be seen, the D^2 distances are shown on the y-axis while the chi-squared quantiles are shown on the x-axis. Of the D^2 distance values, 43 proved to be significant at confidence level $\alpha = 0.001$ (Hair et al. 2019). The maximum D^2 value was 46.29.

Figure 2.23 – Q–Q Plot of Mahalanobis D^2 vs quantiles of χ^2 – Subscale IV.II (frequency)



The number of Guttman errors for each observation was also calculated to identify atypical response patterns. The error average was 14.5 (SD = 13.24); according to the criterion proposed by Zijlstra et al. (2007) and by Hubert and Vandervieren (2008) for asymmetric distributions, the critical value was 42.5, which was exceeded by 187 observations.

Figure 2.24 – Guttman error distribution – Subscale IV.II (frequency)



e Evidence of validity

The Kaiser Meyer Olkin (Kaiser 1970) and Bartlett spherical tests were used to verify data adequacy for factor analysis. The result of the KMO test indicates that the data are adequate for factor analysis (KMO = 0.79). All items of this subscale obtained MSA values greater than or close to 0.7 (i66.1 = 0.87; i66.2 = 0.73; i66.3 = 0.65; i66.4 = 0.81; i66.5 = 0.81; i66.6 = 0.84; i66.7 = 0.81). The result of Bartlett's sphericity test was also significant (χ^2 (21) = 13 756.52; p < 0.001). The results obtained indicated that the data are adequate for submission to factor analysis.

The internal structure is shown in **Figure 2.25**.

Figure 2.26 shows the sedimentation graph with the result of exploratory factor analysis, which also suggests the presence of a single factor. The presence of two factors, however, is supported by eight out of 24 methods (33.33%) (optimal co-ordinates, parallel analysis, Kaiser criterion, Scree (SE), EGA (glasso), EGA (TMFG), VSS complexity 2, Fit_off). It would therefore be appropriate to carry out an in-depth analysis of the internal structure of the scale by means of confirmatory factor analysis.

Figure 2.25 – Measurement module – Subscale IV.II (frequency)

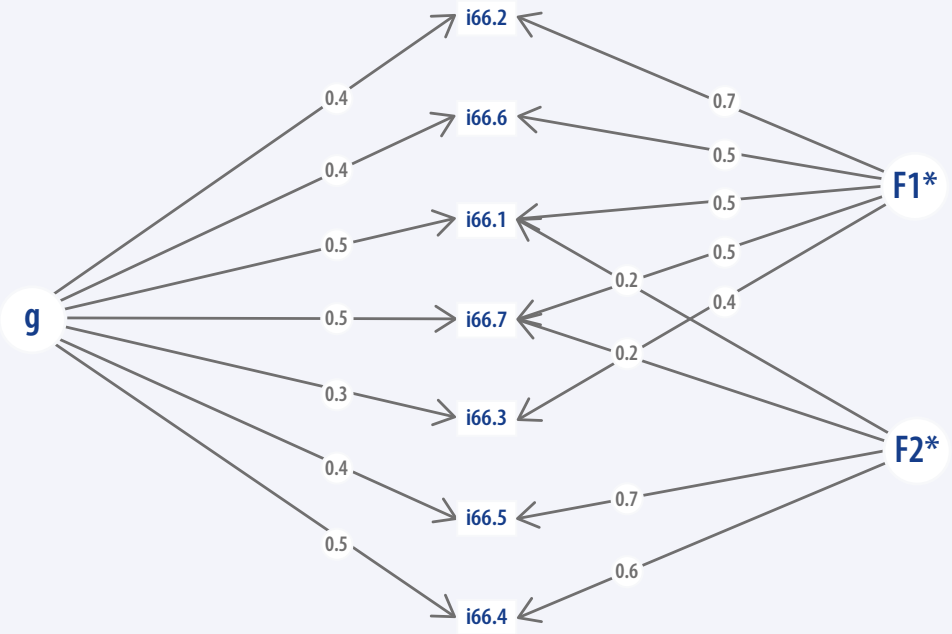
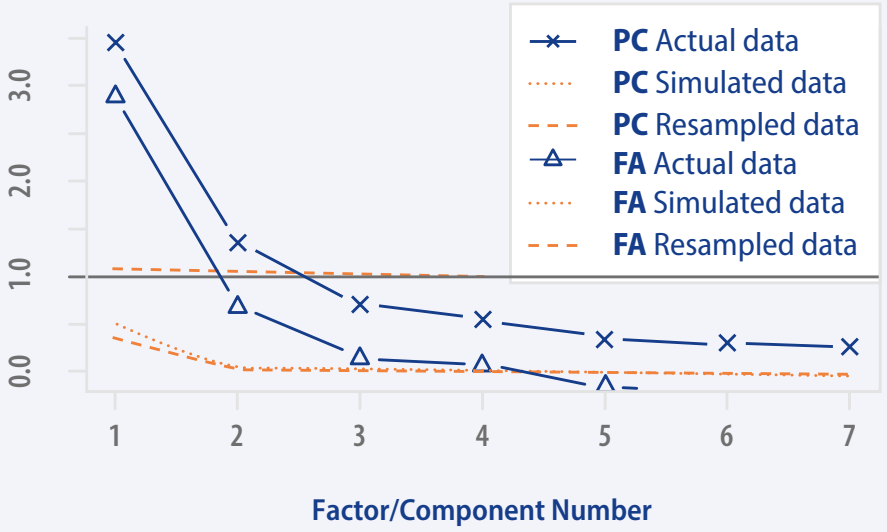


Figure 2.26 – Sedimentation graph – Subscale IV.II (frequency)



Subscale IV.III: influence on teaching practice

a Descriptive analysis

The results of the descriptive analysis of the five items relating to the factors that influence history teaching practice (Subscale IV.III – items 68.1 to 68.5) are shown in **Table 2.30**.

Table 2.30 – Factors that influence practice – Subscale IV.III

Item	Item label	1	2	3	4	5	% 1 to 2	% 4 to 5	Missing
History textbooks	i68.1	179	241	734	1 248	1 733	10.16	72.09	3 732
Exams	i68.2	289	458	1 069	1 330	989	18.07	56.08	3 732
In-service professional development	i68.3	370	537	1 156	1 205	867	21.93	50.11	3 732
Initial teacher training	i68.4	522	727	1 107	992	787	30.21	43.02	3 732
Student needs and interests	i68.5	1 119	413	1 032	843	728	37.05	37.99	3 732

Item	n	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
History textbooks	4 135	4	1.1	4	1	5	4	-1.02	0.37	0.02
Exams	4 135	3.55	1.17	4	1	5	4	-0.53	-0.49	0.02
In-service professional development	4 135	3.4	1.21	4	1	5	4	-0.4	-0.69	0.02
Initial teacher training	4 135	3.19	1.28	3	1	5	4	-0.18	-1	0.02
Student needs and interests	4 135	2.91	1.44	3	1	5	4	-0.04	-1.31	0.02

b Reliability

With respect to the reliability of this set of five items, values below 0.6 were obtained via Cronbach's ordinal alpha (α = 0.57), while values slightly above 0.7 were obtained via McDonald's omega (ω = 0.72) (McDonald 2013; Revelle and Zinbarg 2009). Values below 0.7 are not considered to be acceptable (Kline 1999). **Table 2.31** shows the results of the reliability analysis and the item–total correlations of the scale.

It can be seen that the correlations between three of the items and the total of the scale were below 0.3. These results indicate that the reliability of the scale is low, and it would be advisable to review the drafting of the items and increase their number.

Table 2.31 – Reliability analysis – Subscale IV.III

Item	Alpha if an item is dropped	Item–total correlation
i68.1	0.55	0.26
i68.2	0.56	0.26
i68.3	0.42	0.49
i68.4	0.42	0.48
i68.5	0.59	0.19

c Mokken scale analysis

With respect to item homogeneity, the general scalability coefficient obtained for the five items was H = 0.195 (SE = 0.008). The H scalability values of all items is shown in **Table 2.32**.

Table 2.32 – Homogeneity coefficients – Subscale IV.III

Item	H	SE
i68.1	0.140	(0.011)
i68.2	0.147	(0.011)
i68.3	0.282	(0.009)
i68.4	0.278	(0.009)
i68.5	0.124	(0.012)

All values were below the H = 0.30 threshold. Indications of unidimensionality were therefore not identified and the items are not scalable to H ≥ 0.30.

The automated item selection procedure was then carried out at increasingly homogeneous threshold levels to examine dimensionality. As has been shown, if all items appear to belong to dimension 1, this indicates that the scale is unidimensional within that homogeneity threshold (indicated in the column headings, from 0.1 to 0.5). **Table 2.33** shows the results of the AISP. These results suggest the presence of two dimensions in the scale that will have to be proven by exploratory factor analysis. The first of these involves items 68.3 and 68.4, while the second would involve items 68.1 and 68.2. Item 68.5 does not appear to fit in either of them.

Table 2.33 – MSA – AISP for increasing H thresholds *(t)* – Subscale IV.III

Item	*t* = 0.10	*t* = 0.15	*t* = 0.20	*t* = 0.30	*t* = 0.35	*t* = 0.40	*t* = 0.45	*t* = 0.50
i68.1	0	2	2	2	0	0	0	0
i68.2	1	2	2	2	0	0	0	0
i68.3	1	1	1	1	1	1	1	1
i68.4	1	1	1	1	1	1	1	1
i68.5	1	1	1	0	0	0	0	0

As regards the monotonicity assumption, **Table 2.34** shows the results of the analysis. **Table 2.34** also presents the homogeneity indices of each item. A significant violation (#zsig) of the monotonicity assumption can be observed in dimension 1 but not in dimension 2. That is, the items of dimension 1 appear not to discriminate clearly between respondents with high levels in the construct and those with lower levels.

Table 2.34 – MSA – Monotonicity – Subscale IV.III

Dimension 1 (H = 0.58, SE = 0.014)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i68.3	0.58	40	1	0.03	0.03	0.03	9.00E – 04	2.17	1	18
i68.4	0.58	24	0	0.00	0.00	0.00	0.00E + 00	0.00	0	0

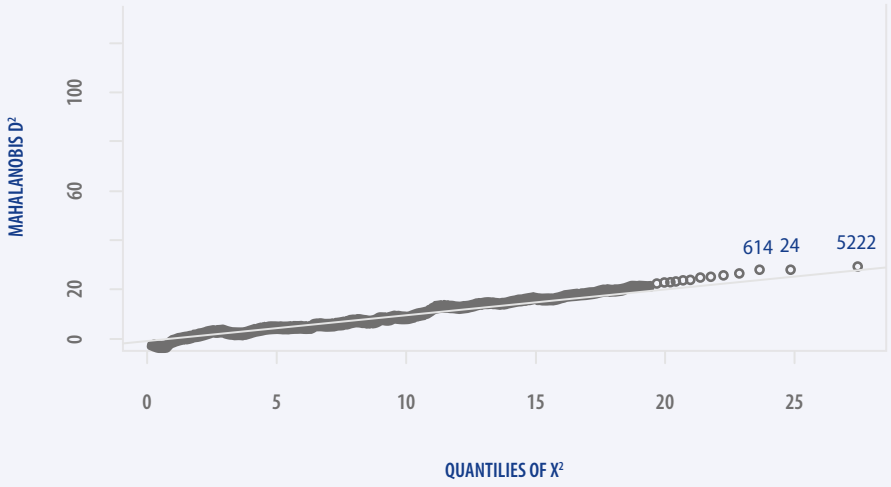
Dimension 2 (H = 0.32, SE = 0.018)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i68.1	0.32	24	0	0	0	0	0	0	0	0
i68.2	0.32	24	0	0	0	0	0	0	0	0

d Multivariate outliers

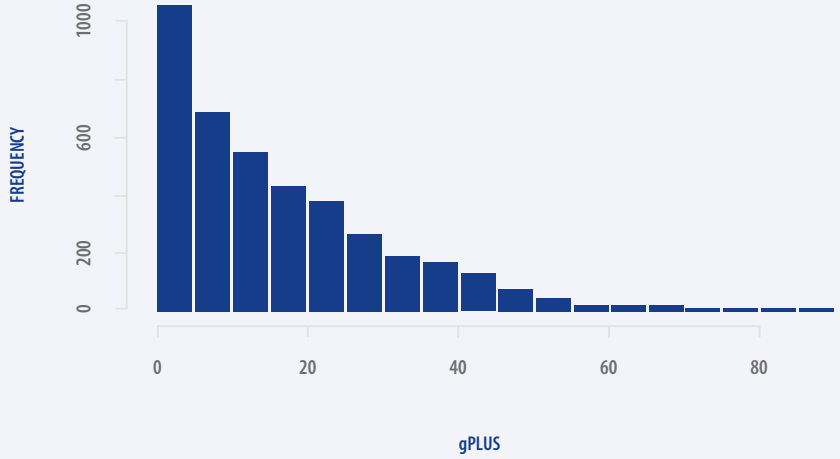
The multivariate outliers were then analysed by means of the Mahalanobis D^2 distances. The results of the analysis are shown in **Figure 2.27**. As can be seen, the D^2 distances are shown on the y-axis while the chi-squared quantiles are shown on the x-axis. Of the figures for the D^2 distances, 12 proved to be significant at confidence level $\alpha = 0.001$ (Hair et al. 2019). The maximum D^2 value was 29.14.

Figure 2.27 – Q–Q Plot of Mahalanobis D^2 vs quantiles of χ^2 – Subscale IV.III



The number of Guttman errors for each observation was also calculated to identify atypical response patterns. The error average was 17.08 (SD = 14.74). According to the criterion proposed by Zijlstra et al. (2007) and by Hubert and Vandervieren (2008) for asymmetric distributions, the critical value was 55. This was exceeded by 77 observations.

Figure 2.28 – Guttman error distribution – Subscale IV.III



e Evidence of validity

To verify data adequacy for factor analysis, the Kaiser Meyer Olkin test (Kaiser 1970) and Bartlett's sphericity tests were used. The result of the KMO test indicates that the data are acceptable for factor analysis (KMO 0.64). All items of this subscale obtained MSA values higher than 0.5 (i68.1 = 0.56; i68.2 = 0.56; i68.3 = 0.58; i68.4 = 0.58; i68.5 = 0.65). The result of Bartlett's sphericity test was also significant (χ^2 (10) = 3452.791; p < 0.001).

The internal structure is shown in **Figure 2.29**.

Figure 2.30 presents the sedimentation graph with the result of exploratory factor analysis, which also suggests the presence of two factors. The presence of two factors is also supported by 10 out of 21 methods (47.62%) (optimal co-ordinates, parallel analysis, Kaiser criterion, VSS complexity 1, BIC, BIC (adjusted), Fit–off, RMSEA, CRMS, BIC).

Figure 2.29 – Measurement module – Subscale IV.III

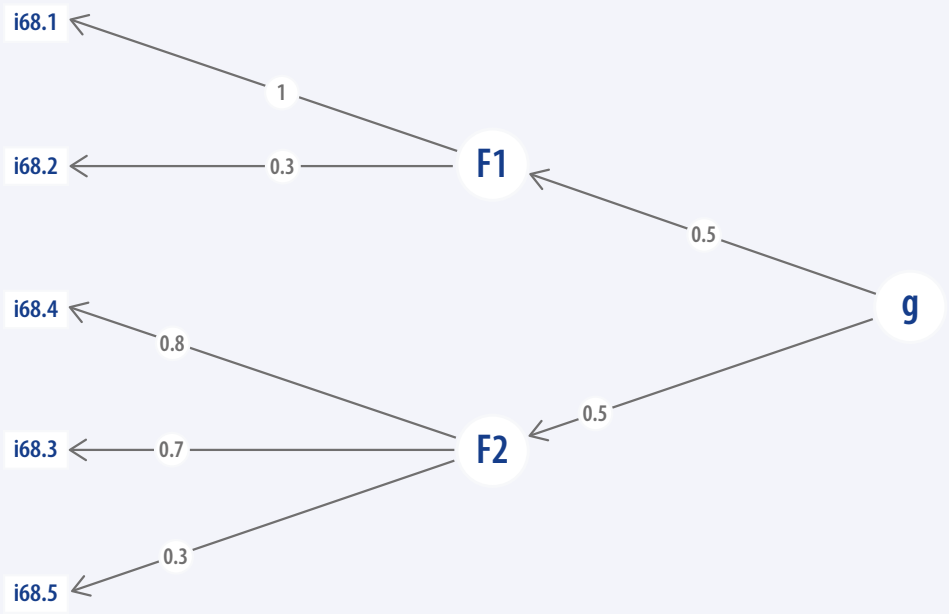
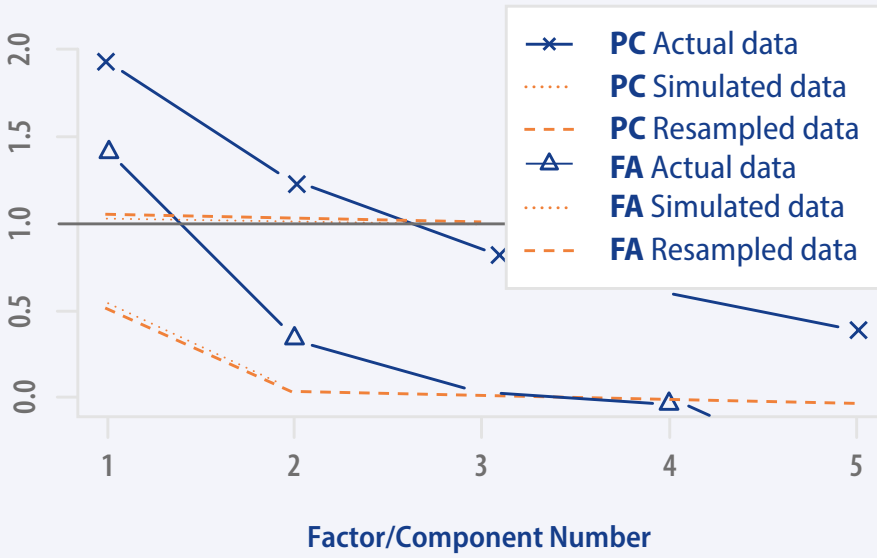


Figure 2.30 – Sedimentation graph – Subscale IV.III



SUBSCALE V – Learning outcomes and assessment

Subscale V.I: teacher aims

a Descriptive analysis

The results of the descriptive analysis of the 14 items relating to the objectives of history teaching (Subscale V – items 70.1 to 70.14) are shown in **Table 2.35**.

Table 2.35 – Frequency of use of different resources – Subscale V.I

Item label	1	2	3	4	5	% 1 to 2	% 4 to 5	No. missing
i70.1	190	543	1 135	1 077	1 155	17.88	54.44	3 767
i70.2	9	54	324	1 126	2 587	1.54	90.56	3 767
i70.3	383	403	625	919	1 770	19.17	65.59	3 767
i70.4	197	385	845	1 274	1 399	14.2	65.2	3 767
i70.5	15	90	352	1 015	2 628	2.56	88.85	3 767
i70.6	12	33	237	887	2 931	1.1	93.12	3 767
i70.7	10	44	310	963	2 773	1.32	91.12	3 767
i70.8	49	181	687	1 385	1 798	5.61	77.63	3 767
i70.9	26	154	607	1 280	2 033	4.39	80.8	3 767
i70.10	17	106	714	1 456	1 807	3	79.59	3 767

Item label	1	2	3	4	5	% 1 to 2	% 4 to 5	No. missing
i70.11	23	88	511	1 224	2 254	2.71	84.83	3 767
i70.12	27	131	622	1 370	1 950	3.85	80.98	3 767
i70.13	56	191	656	1 234	1 963	6.02	77.98	3 767
i70.14	28	97	429	950	2 596	3.05	86.49	3 767

Item	n	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
i70.1	4 100	3.6	1.16	4	1	5	4	-0.4	-0.73	0.02
i70.2	4 100	4.52	0.72	5	1	5	4	-1.5	2.09	0.01
i70.3	4 100	3.8	1.33	4	1	5	4	-0.83	-0.55	0.02
i70.4	4 100	3.8	1.15	4	1	5	4	-0.75	-0.26	0.02
i70.5	4 100	4.5	0.77	5	1	5	4	-1.6	2.32	0.01
i70.6	4 100	4.63	0.66	5	1	5	4	-1.97	4.27	0.01
i70.7	4 100	4.57	0.7	5	1	5	4	-1.69	2.74	0.01
i70.8	4 100	4.15	0.93	4	1	5	4	-0.98	0.49	0.01
i70.9	4 100	4.25	0.89	4	1	5	4	-1.06	0.55	0.01
i70.10	4 100	4.2	0.85	4	1	5	4	-0.82	0.11	0.01
i70.11	4 100	4.37	0.82	5	1	5	4	-1.24	1.18	0.01
i70.12	4 100	4.24	0.87	4	1	5	4	-1.02	0.58	0.01
i70.13	4 100	4.18	0.96	4	1	5	4	-1.07	0.58	0.01
i70.14	4 100	4.46	0.83	5	1	5	4	-1.57	2.14	0.01

b Reliability

With respect to the reliability of this set of 14 items, values higher than 0.8 were obtained by means of Cronbach's ordinal alpha (α = 0.92) and of McDonald's omega (ω = 0.94) (McDonald 2013; Revelle and Zinbarg 2009). Values higher than 0.9 are considered to be excellent (Kline 1999). **Table 2.36** presents the results of the reliability analysis and the item–total correlations of the scale.

It can be seen that all item–total correlations were greater than 0.3.

Table 2.36 – Reliability analysis – Subscale V.I

Item	Alpha if an item is dropped	Item–total correlation
i70.1	0.92	0.37
i70.2	0.91	0.69
i70.3	0.92	0.38
i70.4	0.92	0.50
i70.5	0.91	0.69
i70.6	0.91	0.75
i70.7	0.91	0.75
i70.8	0.91	0.71
i70.9	0.91	0.71
i70.10	0.91	0.70
i70.11	0.91	0.70
i70.12	0.91	0.73
i70.13	0.91	0.68
i70.14	0.91	0.69

c Mokken scale analysis

With respect to item homogeneity, the general scalability coefficient obtained for the 14 items was H = 0.372 (SE = 0.008). The H scalability values of all items are shown in **Table 2.37**.

All values exceeded threshold H = 0.30, except for items 70.1 and 70.3.

The automated item selection procedure was then carried out at increasingly homogeneous threshold levels to examine dimensionality. As shown above, if all items appear to belong to dimension 1, this indicates that the scale is unidimensional in that homogeneity threshold (indicated in the column headings, from 0.1 to 0.5). **Table 2.38** shows the results of the AISP. These results suggest the presence of two dimensions in the scale that will have to be proven by means of exploratory factor analysis. The first of these consists of all items except for 70.1 and 70.3, which will form the second dimension.

Table 2.37 – Homogeneity coefficients – Subscale V.I

Item	H	SE
i70.1	0.245	(0.011)
i70.2	0.407	(0.012)
i70.3	0.239	(0.010)
i70.4	0.314	(0.011)
i70.5	0.395	(0.011)
i70.6	0.435	(0.012)
i70.7	0.435	(0.011)
i70.8	0.421	(0.009)
i70.9	0.422	(0.010)
i70.10	0.416	(0.010)
i70.11	0.404	(0.011)
i70.12	0.431	(0.010)
i70.13	0.400	(0.010)
i70.14	0.400	(0.010)

Table 2.38 – MSA – AISP for increasing H thresholds *(t)* – Subscale V.I

Item	*t* = 0.10	*t* = 0.15	*t* = 0.20	*t* = 0.30	*t* = 0.35	*t* = 0.40	*t* = 0.45	*t* = 0.50
i70.1	1	1	1	2	2	2	0	0
i70.2	1	1	1	1	1	1	1	0
i70.3	1	1	1	2	2	2	0	0
i70.4	1	1	1	1	0	0	0	0
i70.5	1	1	1	1	1	1	1	1
i70.6	1	1	1	1	1	1	1	1
i70.7	1	1	1	1	1	1	1	1
i70.8	1	1	1	1	1	1	1	1
i70.9	1	1	1	1	1	1	1	0
i70.10	1	1	1	1	1	1	1	1
i70.11	1	1	1	1	1	1	1	1
i70.12	1	1	1	1	1	1	1	1
i70.13	1	1	1	1	1	1	1	2
i70.14	1	1	1	1	1	1	1	2

With respect to the monotonicity assumption, **Table 2.39** shows the results of the analysis. **Table 2.39** also presents the homogeneity indices of each item. No significant violations (#zsig) of the monotonicity assumption are observed for any of the items of Subscale V.I: teacher aims. That is, all items appear to discriminate clearly between respondents with high levels in the construct and those with lower levels.

Table 2.39 – MSA – Monotonicity – Subscale V.I

Dimension 1 (H = 0.45, SE = 0.009)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i70.2	0.43	87	0	0	0	0	0	0	0	0
i70.4	0.31	112	0	0	0	0	0	0	0	0
i70.5	0.46	76	0	0	0	0	0	0	0	0
i70.6	0.48	90	0	0	0	0	0	0	0	0
i70.7	0.50	75	0	0	0	0	0	0	0	0
i70.8	0.48	98	0	0	0	0	0	0	0	0
i70.9	0.45	92	0	0	0	0	0	0	0	0
i70.10	0.47	83	0	0	0	0	0	0	0	0
i70.11	0.49	96	0	0	0	0	0	0	0	0
i70.12	0.50	96	0	0	0	0	0	0	0	0
i70.13	0.46	105	0	0	0	0	0	0	0	0
i70.14	0.47	71	0	0	0	0	0	0	0	0

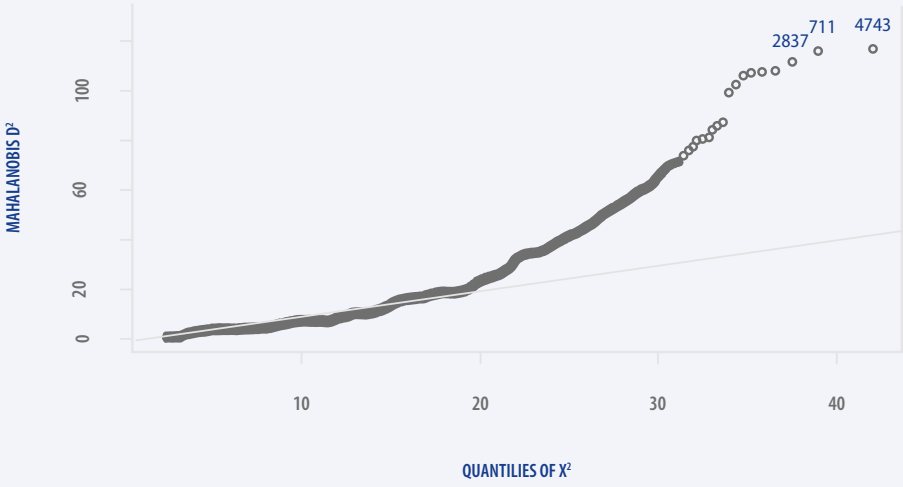
Dimension 2 (H = 0.43, SE = 0.016)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i70.1	0.43	24	0	0	0	0	0	0	0	0
i70.3	0.43	24	0	0	0	0	0	0	0	0

d Multivariate outliers

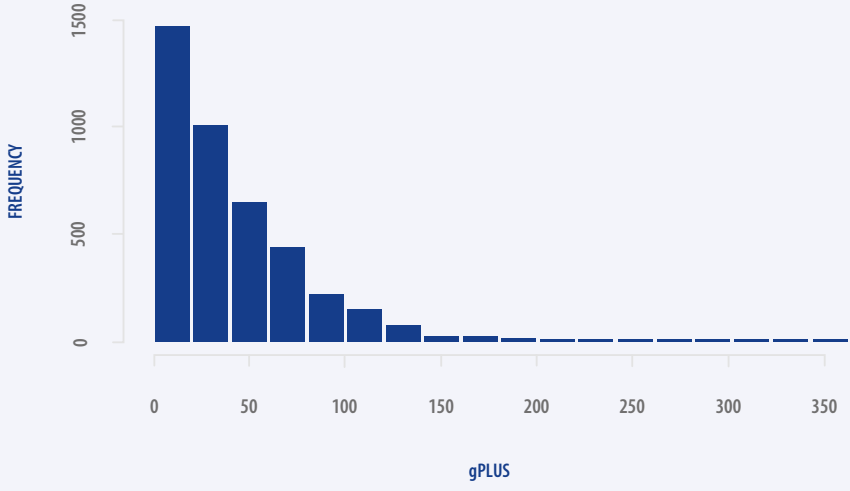
The multivariate outliers were then analysed by means of the Mahalanobis D^2 distances. The results of the analysis are shown in **Figure 2.31**. As can be seen, the D^2 distances are shown on the y-axis while the chi-squared quantiles are shown on the x-axis. Of the D^2 distance values, 218 were significant at confidence level $\alpha = 0.001$ (Hair et al. 2019). The maximum D^2 value was 117.13.

Figure 2.31 – Q–Q Plot of Mahalanobis D^2 vs quantiles of χ^2 – Subscale V.I



The number of Guttman errors for each observation was also calculated to identify atypical response patterns. The error average was 41.78 (SD = 39.60). According to the criterion proposed by Zijlstra et al. (2007) and Hubert and Vandervieren (2008) for asymmetric distributions, the critical value was 126.5, which was exceeded by 159 observations.

Figure 2.32 – Guttman error distribution – Subscale V.I



e Evidence of validity

Kaiser Meyer Olkin tests (Kaiser 1970) and Bartlett's sphericity test were used to verify data adequacy for factor analysis. The result of the KMO test indicates that the data are acceptable for factor analysis (KMO = 0.92). All items of this subscale obtained MSA values higher than 0.9, except for items 70.3 (MSA = 0.74) and 70.4 (MSA = 0.85). The result of Bartlett's sphericity test was also significant (χ^2 (91) = 38 359.77; p < 0.001).

The internal structure is shown in **Figure 2.33**.

Figure 2.34 shows the sedimentation graph with the result of exploratory factor analysis, which also suggests the presence of two factors. However, the presence of three factors is supported by six methods out of 27 (22.22%) (CNG, optimal co-ordinates, parallel analysis, Kaiser criterion, Scree (SE), EGA (glasso)). It would therefore be appropriate to carry out a deeper analysis of the internal structure of the scale, submitting the respective data to a confirmatory factor analysis.

Figure 2.33 – Measurement module – Subscale V.I

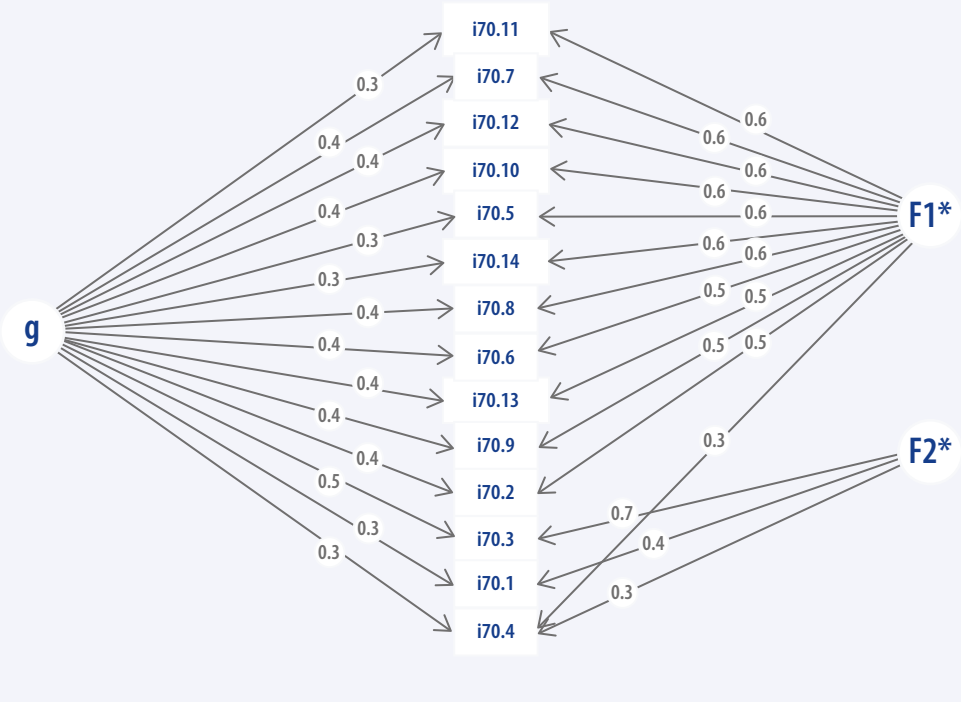
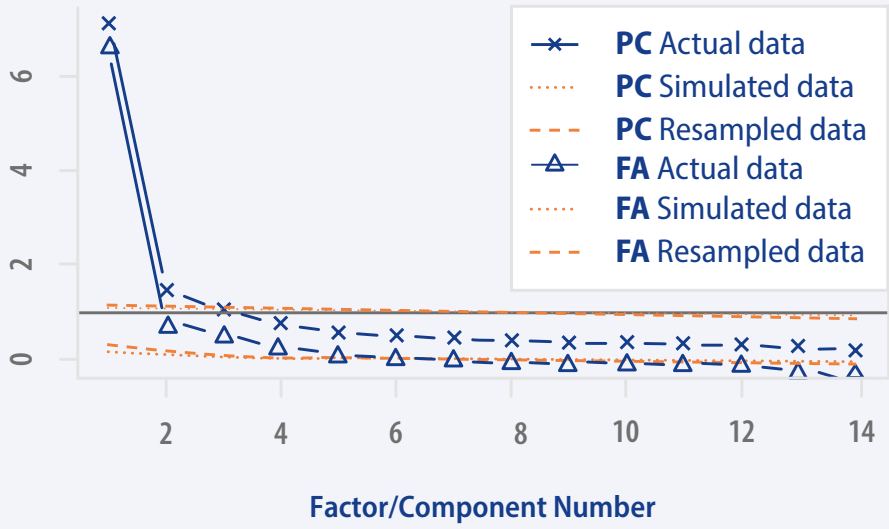


Figure 2.34 – Sedimentation graph – Subscale V.I



Subscale V.II: methods of assessment

a Descriptive analysis

The results of the descriptive analysis of the 10 items relating to the assessment instruments (Subscale V – items 73.1 to 73.10) are shown in **Table 2.40**.

Table 2.40 – Factors that influence practice – Subscale V.II

Item label	1	2	3	4	5	% 1 to 2	% 4 to 5	No. missing
i73.1	107	288	1 004	1 499	1 157	9.74	65.5	3 812
i73.2	76	261	833	1 390	1 495	8.31	71.15	3 812
i73.3	118	379	1 039	1 434	1 085	12.26	62.12	3 812
i73.4	193	493	1 076	1 356	937	16.92	56.55	3 812
i73.5	481	861	1 210	930	573	33.09	37.07	3 812
i73.6	184	569	1 192	1 202	908	18.57	52.03	3 812
i73.7	174	451	1 085	1 363	982	15.41	57.83	3 812
i73.8	74	207	538	1 217	2 019	6.93	79.8	3 812
i73.9	145	490	1 280	1 345	795	15.66	52.77	3 812
i73.10	250	494	1 203	1 236	872	18.37	51.99	3 812

Item	n	Mean	SD	Median	Min	Max	Range	Skew	Kurtosis	SE
i70.1	4 055	3.82	1.01	4	1	5	4	-0.65	-0.02	0.02
i70.2	4 055	3.98	1	4	1	5	4	-0.79	0.05	0.02
i70.3	4 055	3.74	1.04	4	1	5	4	-0.57	-0.27	0.02
i70.4	4 055	3.58	1.11	4	1	5	4	-0.48	-0.48	0.02
i70.5	4 055	3.06	1.22	3	1	5	4	-0.04	-0.9	0.02
i70.6	4 055	3.51	1.12	4	1	5	4	-0.34	-0.65	0.02
i70.7	4 055	3.62	1.1	4	1	5	4	-0.51	-0.41	0.02
i70.8	4 055	4.21	0.98	4	1	5	4	-1.22	0.97	0.02
i70.9	4 055	3.53	1.05	4	1	5	4	-0.35	-0.44	0.02
i70.10	4 055	3.49	1.14	4	1	5	4	-0.41	-0.54	0.02

b Reliability

With respect to the reliability of this set of 10 items, values higher than 0.8 were obtained both by Cronbach's ordinal alpha (α = 0.87) and McDonald's omega (ω = 0.9) (McDonald 2013; Revelle and Zinbarg 2009). Values higher than 0.8 are considered to be good (Kline 1999). **Table 2.41** shows the results of the reliability analysis and the item–total correlations of the scale.

It can be seen that all item–total correlations were greater than 0.3.

Table 2.41 – Reliability analysis – Subscale V.II

Item	Alpha if an item is dropped	Item–total correlation
i73.1	0.86	0.57
i73.2	0.86	0.52
i73.3	0.86	0.56
i73.4	0.85	0.66
i73.5	0.86	0.57
i73.6	0.86	0.56
i73.7	0.85	0.59
i73.8	0.87	0.43
i73.9	0.85	0.68
i73.10	0.85	0.68

c Mokken scale analysis

With respect to item homogeneity, the general scalability coefficient obtained for the 10 items was H = 0.37 (SE = 0.007). The H scalability values of all items are shown in **Table 2.42**.

All values were greater than the threshold H = 0.30, except for item 73.8.

The automated item selection procedure was then carried out at increasing threshold levels to examine dimensionality. As has been shown, if all items appear to belong to dimension 1, this indicates that the scale is unidimensional within that homogeneity threshold (indicated in the column headings, from 0.1 to 0.5).

Table 2.43 shows the results of the AISP. These results suggest the presence of a dimension in the scale that will have to be proven by means of exploratory factor analysis. Once again, item 73.8 appears to remain outside the unidimensional scale.

Table 2.42 – Homogeneity coefficients – Subscale V.II

Item	H	SE
i73.1	0.363	(0.010)
i73.2	0.325	(0.010)
i73.3	0.353	(0.011)
i73.4	0.414	(0.009)
i73.5	0.381	(0.010)
i73.6	0.357	(0.010)
i73.7	0.369	(0.009)
i73.8	0.262	(0.010)
i73.9	0.434	(0.009)
i73.10	0.426	(0.009)

Table 2.43 – MSA – AISP for increasing H thresholds *(t)* – Subscale V.II

Item	*t* = 0.10	*t* = 0.15	*t* = 0.20	*t* = 0.30	*t* = 0.35	*t* = 0.40	*t* = 0.45	*t* = 0.50
i73.1	1	1	1	1	1	2	2	0
i73.2	1	1	1	1	2	3	0	0
i73.3	1	1	1	1	1	2	0	0
i73.4	1	1	1	1	1	1	1	2
i73.5	1	1	1	1	1	1	1	2
i73.6	1	1	1	1	1	1	0	0
i73.7	1	1	1	1	1	2	2	0
i73.8	1	1	1	0	2	3	0	0
i73.9	1	1	1	1	1	1	1	1
i73.10	1	1	1	1	1	1	1	1

With respect to the monotonicity assumption, **Table 2.44** shows the results of the analysis. **Table 2.44** also shows the homogeneity indices of each item. Neither significant (#zsig) nor non-significant (#vi) violations of the monotonicity assumption are observed for any items in Subscale V.II: methods for assessment. That is, all items appear to discriminate well between respondents with high levels in the construct and those with lower levels.

Table 2.44 – MSA – Monotonicity – Subscale V.II

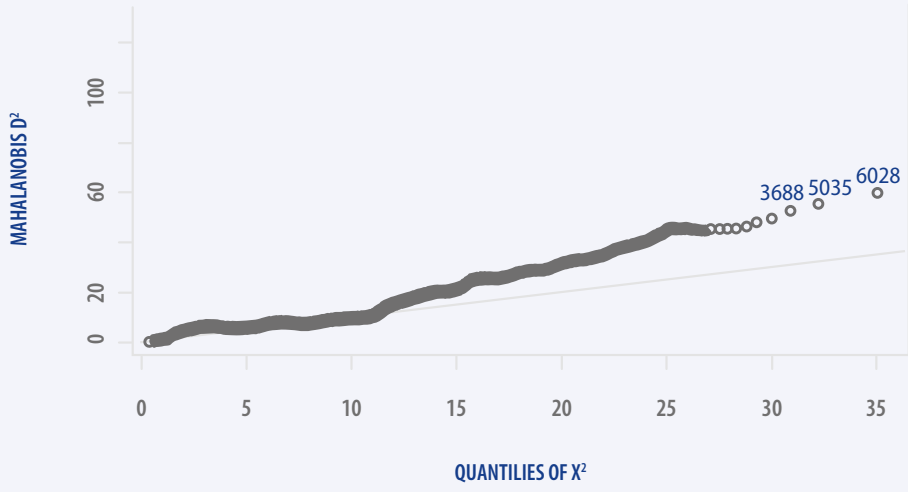
Dimension 1 (H = 0.37, SE = 0.008)

Item	Item	#ac	#vi	#vi/#ac	maxvi	sum	sum/#ac	zmax	#zsig	crit
i73.1	0.38	112	0	0	0	0	0	0	0	0
i73.2	0.32	112	0	0	0	0	0	0	0	0
i73.3	0.37	112	0	0	0	0	0	0	0	0
i73.4	0.43	105	0	0	0	0	0	0	0	0
i73.5	0.4	112	0	0	0	0	0	0	0	0
i73.6	0.37	105	0	0	0	0	0	0	0	0
i73.7	0.38	112	0	0	0	0	0	0	0	0
i73.9	0.45	105	0	0	0	0	0	0	0	0
i73.10	0.44	112	0	0	0	0	0	0	0	0

d Multivariate outliers

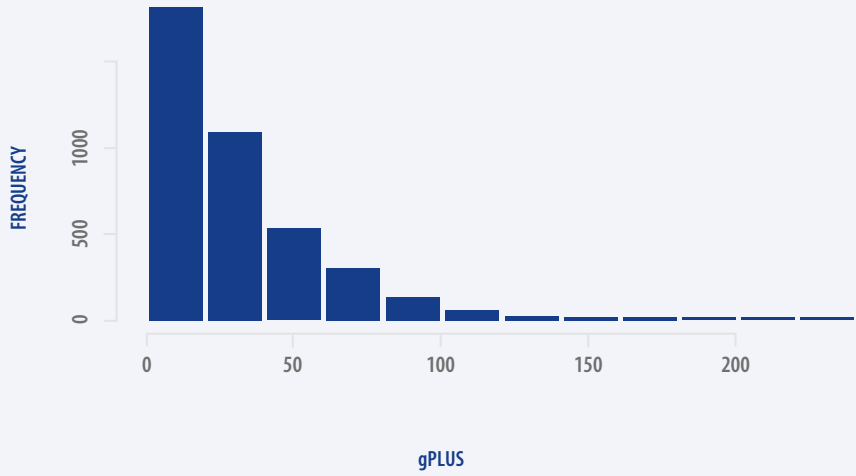
The multivariate outliers were then analysed by means of the Mahalanobis D^2 distances. The results of the analysis are shown in **Figure 2.35**. As can be seen, the D^2 distances are shown on the y-axis while the chi-squared quantiles are shown on the x-axis. Of the D^2 distance values, 108 were significant at confidence level $\alpha = 0.001$ (Hair et al. 2019). The maximum D^2 value was 60.03.

Figure 2.35 – Q–Q Plot of Mahalanobis D^2 vs quantiles of χ^2 – Subscale V.II



The number of Guttman errors for each observation was also calculated to identify atypical response patterns. The error average was 30.32 (SD = 27.46). According to the criterion proposed by Zijlstra et al. (2007) and Hubert and Vandervieren (2008) for asymmetric distributions, the critical value was 92.5, which was exceeded by 148 observations.

Figure 2.36 – Guttman error distribution – Subscale V.II



e Evidence of validity

The Kaiser Meyer Olkin tests (Kaiser 1970) and Bartlett spherical tests were used to verify the conformity of the data for factor analysis. The result of the KMO test indicates that the data are acceptable for factor analysis (KMO 0.89). All items of this subscale obtained MSA values between 0.84 (item 73.8) and 0.92 (item 73.5). The result of Bartlett's sphericity test was also significant (χ^2 (45) = 16781.77; p < 0.001).

The internal structure is shown in **Figure 2.37**.

Figure 2.38 shows the sedimentation graph with the result of exploratory factor analysis, which suggests the presence of one factor. The presence of one factor is supported by nine out of 27 methods (29.63%) (optimal co-ordinates, acceleration factor, Scree (SE), Scree (R2), EGA (glasso), EGA (TMFG), VSS complexity 1, Velicer's MAP, TLI).

Figure 2.37 – Measurement module – Subscale V.II

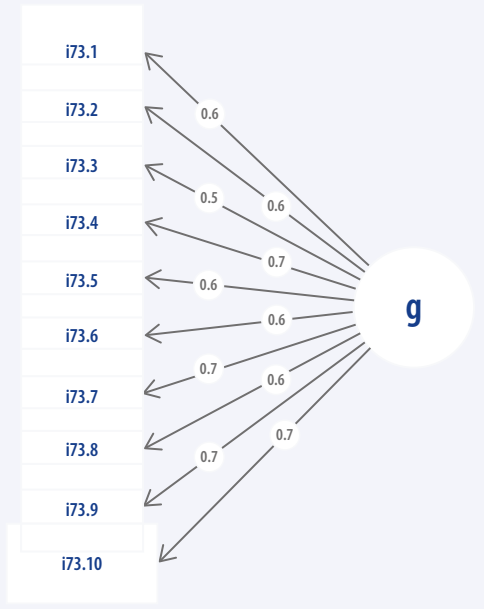
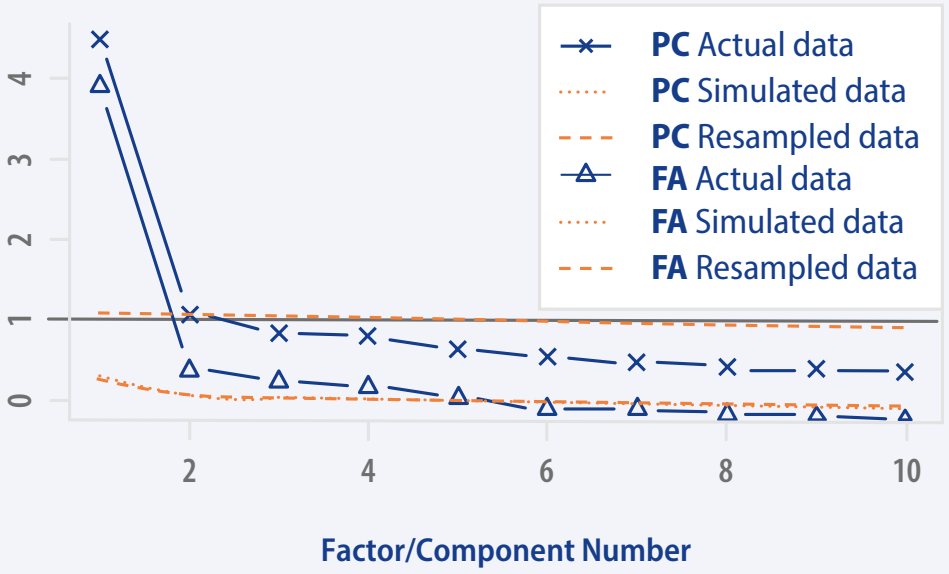


Figure 2.38 – Sedimentation graph – Subscale V.II



References

Ark L. A. (2007), "Mokken scale analysis", in *Journal of Statistical Software* Vol. 20, No. 11, pp. 1-19.

DeSimone J. A. et al. (2018), "The differential impacts of two forms of insufficient effort responding", *Applied Psychology*, Vol. 67, No. 2, pp. 309-38. doi: 10.1111/apps.12117.

Hair J. F. et al. (2019), *Multivariate data analysis* (8th edn), Cengage Learning, Andover.

Hubert M. and Vandervieren E. (2008), "An adjusted boxplot for skewed distributions", *Computational Statistics and Data Analysis* Vol. 52, No. 12, pp. 5186-5201. https://doi.org/10.1016/j.csda.2007.11.008.

Kaiser H. F. (1970), "A second generation little jiffy", *Psychometrika*, 35, pp. 401-415 https://doi.org/10.1007/BF02291817.

Kaiser H. F. and Rice J. (1974), "Little Jiffy Mark IV", *Educational and Psychological Measurement* Vol. 34, No. 1, pp. 111-117, https://doi.org/10.1177/001316447403400115.

Kline P. (1999), *The handbook of psychological testing* (2nd edn), Routledge, Abingdon.

Loevinger J. (1948), "The technique of homogeneous tests compared with some aspects of 'scale analysis' and factor analysis", *Psychological Bulletin*, 45, pp. 507-530. https://doi.org/10.1037/h0055827.

Lorenzo-Seva U. and Ferrando P. J. (2021), "Not positive definite correlation matrices in exploratory item factor analysis: Causes, consequences and a proposed solution", *Structural Equation Modeling: A Multidisciplinary Journal* Vol. 28, No. 1, pp. 138-147. http://doi.org/10.1080/10705511.2020.1735393.

McDonald R. P. (2013), *Test theory: a unified treatment,* Larence Erlbaum, Psychology Press Mahwah NJ.

Mokken R. J. (1971), *A theory and procedure of scale analysis*, Mouton, The Hague.

Molenaar I. W. and Sjitsma K. (1984), "Internal consistency and reliability in Mokken's nonparametric item response model", *Tijdschrift voor Onderwijsresearch* Vol. 9, pp. 257-68.

Revelle W. and Zinbarg R. E. (2009), "Coefficients alpha, beta, omega and the glb: comments on Sijtsma", *Psychometrika* Vol. 74, No. 1, pp. 145-54.

Stochl J. et al. (2012), "Mokken scale analysis of mental health and well-being questionnaire item responses: a non-parametric IRT method in empirical research for applied health researchers", *BMC Medical Research Methodology* Vol. 12, 74. https://doi.org/10.1186/1471-2288-12-74.

Zijlstra W. P., Van Der Ark L. A. and Sijtsma K. (2007), "Outlier detection in test and questionnaire data", *Multivariate Behavioral Research* Vol. 42, No. 3, pp. 531-555. https://doi.org/10.1080/00273170701384340.

This analysis was conducted by Jairo Rodríguez Medina (University of Valladolid, Spain) and Cosme Jesús Gómez Carrasco (University of Murcia, Spain).

ITEM 3

TABLES WITH MEAN VALUES OF RESPONSES TO SELECTED QUESTIONS IN THE TES

Item 3

Tables with mean values of responses to selected questions in the TES

Table 3.1 – Frequency of use of educational resources, as indicated by TES respondents, ranging from 1 (never) to 5 (every lesson)

Member state	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
ALB	2.70	2.85	2.54	3.01	2.65	2.79	3.35	3.07	4.73	3.34	3.33	3.25	2.74	2.66	3.8	2.03	3.5
AND	3.29	3.00	3.00	3.00	2.86	2.57	3.71	3.57	3.71	1.57	2.86	3.43	3.71	4.29	2.71	2.43	3.86
ARM	3.09	2.78	3.32	2.98	3.19	2.62	3.19	3.14	4.62	3.28	3.33	3.02	2.51	3.11	3.63	2.32	3.26
CYP	1.83	3.14	2.94	3.07	2.12	1.75	2.62	2.51	4.12	2.35	3.65	3.16	2.38	3.38	4.31	1.67	3.28
FRA	1.86	3.19	3.15	2.41	2.46	1.98	2.71	2.56	3.74	1.93	3.09	3.16	2.26	2.68	2.17	1.56	3.21
GEO	2.98	3.21	3.27	2.99	2.83	2.57	3.42	3.21	4.7	3.08	3.79	3.26	2.8	3.34	2.73	1.96	3.12
GRC	1.97	3.06	3.1	3.02	2.27	1.78	2.6	3.06	4.24	2.28	3.01	2.94	2.26	3.65	4.08	1.67	3.53
IRL	2.46	2.92	3.41	2.63	2.0	2.45	2.96	2.76	3.95	2.58	3.92	3.46	2.69	2.97	4.12	1.53	3.41
LUX	2.15	2.8	3.37	2.16	2.0	2.12	2.57	2.57	3.39	2.27	3.52	3.04	2.44	2.9	3.19	1.68	2.8
MLT	1.78	3.27	3.6	2.65	2.17	2.43	3.4	3.18	2.85	2.28	3.7	3.43	2.58	3.48	4.17	1.83	3.35
MKD	3.13	3.13	3.21	3.29	2.73	2.6	3.25	3.06	4.57	2.97	3.3	3.58	2.91	3.16	3.82	1.88	3.49
PRT	2.84	3.15	3.51	3.05	2.52	2.61	3.19	3.01	4.20	2.38	3.23	3.07	2.22	3.16	3.71	1.96	3.45
SRB	2.71	2.95	3.24	2.98	2.8	2.45	3.16	2.9	4.59	2.7	3.22	3.18	2.69	3.24	3.66	1.71	3.49
SVN	2.61	2.93	3.62	2.92	2.78	2.68	3.13	3.16	3.97	2.57	3.09	3.17	2.76	3.45	3.27	1.99	3.34
ESP	2.55	3.22	3.45	2.53	2.55	2.7	3.18	2.89	3.56	2.47	3.12	2.97	2.6	3.23	3.67	1.77	3.13
TUR	3.34	3.1	3.35	3.0	2.71	2.49	3.22	2.95	4.43	2.69	2.99	3.66	2.8	3.06	4.05	2.3	3.7

- A – Apps for smartphones and tablets with historical content
- B – Artefacts (e.g., paintings, architecture, sculptures, contemporary art)
- C – Cinema and documentaries with historical themes
- D – Historiographical bibliography
- E – Literature (e.g., historical novels, graphic novels)
- F – Local and regional festivals and traditions related to historical events
- G – Local cultural heritage (e.g., costumes, food traditions, celebrations)
- H – Museums and other places of heritage interpretation
- I – History textbooks
- J – Oral sources (interviews with grandparents, relatives, neighbours, etc.)
- K – Primary documentary sources
- L – Printed or digital press (newspapers and magazines)

M Reports on historical topics in popular magazines

N Search engines and websites with historical content not necessarily validated by the education authorities

O Teacher notes

P Video games

Q Cinema and documentaries with historical themes

Table 3.2 – Views of TES respondents on the history textbooks available in their countries, ranging from 1 (strongly disagree) to 5 (strongly agree)

Member state	A	B	C	D	E	F	G	H	I	J	K	L
ALB	3.88	2.38	3.94	3.52	3.80	3.04	2.69	3.75	3.16	2.91	2.71	3.85
AND	3.50	2.50	3.00	1.83	3.50	2.50	2.17	2.17	3.17	2.17	2.50	3.50
ARM	3.37	2.35	3.15	3.21	3.02	2.39	1.98	2.99	1.87	1.94	2.48	2.82
CYP	2.48	3.56	3.05	3.15	2.33	2.07	1.85	2.15	1.48	1.43	3.03	2.19
FRA	3.39	2.62	2.97	2.80	2.96	2.33	1.74	2.48	2.07	2.05	2.37	3.05
GEO	3.76	1.90	3.46	2.11	3.27	2.65	2.11	3.52	2.69	2.16	2.18	3.65
GRC	2.52	3.39	2.93	3.47	2.21	1.97	1.74	2.0	1.4	1.41	2.91	2.22
IRL	3.32	2.64	3.38	3.49	2.89	2.97	2.41	2.57	1.6	1.87	2.39	3.11
LUX	3.18	2.89	3.09	3.01	2.99	2.44	2.22	2.71	2.24	2.29	2.15	3.06
MLT	2.93	2.71	3.25	3.24	3.02	2.40	1.91	2.62	1.75	2.02	2.18	2.8
MKD	3.24	2.85	2.89	2.56	3.23	2.73	2.56	3.48	2.59	2.61	2.94	3.23
PRT	3.69	2.28	3.46	2.43	3.28	2.52	2.09	2.9	2.29	2.43	2.36	3.35
SRB	3.74	2.76	3.61	2.95	3.28	2.93	2.53	3.35	2.96	2.87	2.37	3.52
SVN	4.04	2.76	3.98	3.17	3.55	3.13	2.83	3.08	2.38	2.39	2.48	3.65
ESP	2.98	2.88	2.86	3.35	2.43	2.17	1.75	2.23	2.03	1.91	2.72	2.68
TUR	3.12	3.04	3.1	3.86	3.05	2.78	2.36	2.89	2.67	2.41	2.52	2.87

A The history textbooks provide the necessary material and activities for the development of historical thinking concepts and skills related to how we learn about the past

B The history textbooks set constraints on the way I teach history

C The history textbooks use unbiased language

D The history textbooks present a nation-centred narrative

E The history textbooks present multiple perspectives

F Gender history has an appropriate place in the history textbooks

G The history of childhood has an appropriate place in the history textbooks

H Different ethnic, linguistic, religious and socio-cultural groups are presented adequately in the history textbooks

I Roma and Travellers are presented adequately in the history textbooks

J Different sexual/gender minorities are presented adequately in the history textbooks

K The historical information provided in the history textbooks is outdated

L The methods that are used in history textbooks are suited to the needs of students

Table 3.3 – Importance of different fields in history teaching, as indicated by TES respondents, ranging from 1 (not important at all) to 5 (very important)

Member state	A	B	C	D	E	F	G
ALB	3.41	3.75	3.61	3.02	3.59	3.82	3.89
AND	3.67	4.17	3.33	2.67	2.83	3.50	3.50
ARM	4.16	4.02	4.64	2.71	3.37	3.51	3.92
CYP	3.73	4.15	4.05	3.41	3.68	3.71	3.43
FRA	3.54	4.06	3.82	2.89	3.21	3.43	3.19
GEO	3.85	4.02	4.38	3.48	3.81	4.03	3.94
GRC	3.80	4.32	3.77	3.30	3.66	3.76	3.37
IRL	2.81	4.18	4.21	3.12	3.35	3.48	2.75
LUX	2.87	4.13	3.94	2.47	3.32	3.59	2.85
MLT	3.38	4.35	3.92	3.12	3.19	3.29	3.31
MKD	3.38	3.55	3.96	2.90	3.60	3.57	3.50
PRT	3.98	4.51	4.18	3.04	3.25	3.42	3.18
SRB	3.64	4.20	4.42	3.03	3.62	3.67	3.38
SVN	3.52	4.13	3.92	3.22	3.29	3.43	3.41
ESP	4.02	4.46	3.63	3.50	3.46	3.68	3.37
TUR	3.85	4.41	4.29	3.03	3.71	3.94	4.09

A Art history

B Social and economic history

C Political and military history

D Gender history

E History of minorities and cultures

F Migration history

G Environmental history

Table 3.4 – Frequency of use of different fields in history teaching, as indicated by TES respondents, ranging from 1 (never) to 5 (every lesson)

Member state	A	B	C	D	E	F	G
ALB	2.73	3.27	3.32	2.35	2.93	2.60	2.71
AND	2.67	4.00	3.33	2.33	3.00	3.00	3.33
ARM	4.04	3.97	4.56	2.59	3.07	3.03	3.39
CYP	2.74	3.81	4.37	1.97	2.52	2.76	1.91
FRA	3.21	3.85	3.94	2.36	2.58	2.77	2.52
GEO	2.60	3.05	3.41	2.45	2.69	2.72	2.63
GRC	2.79	3.60	4.05	1.98	2.34	2.46	1.87
IRL	2.16	3.95	4.29	2.39	2.45	2.58	1.82
LUX	2.36	3.97	4.00	2.12	3.05	2.92	2.14

Member state	A	B	C	D	E	F	G
MLT	2.83	4.00	4.15	2.19	2.21	2.64	2.64
MKD	3.14	3.35	4.12	2.71	3.52	3.03	2.80
PRT	3.92	4.42	4.16	2.82	2.96	3.02	2.65
SRB	3.49	4.17	4.55	2.70	3.26	3.31	2.72
SVN	3.16	4.00	4.32	2.84	2.91	2.97	2.56
ESP	3.46	4.07	3.90	3.04	2.78	2.84	2.55
TUR	3.02	3.64	3.91	2.91	3.63	3.00	3.09

A Art history

B Social and economic history

C Political and military history

D Gender history

E History of minorities and cultures

F Migration history

G Environmental history

Table 3.5 – Emphasis given to geographical scales of history, as indicated by TES respondents, ranging from 1 (least relevant) to 5 (most relevant)

Member state	A	B	C	D	E
ALB	2.24	3.26	2.72	3.31	3.47
AND	2.00	3.17	1.83	4.50	3.50
ARM	2.30	2.68	2.86	3.46	3.70
CYP	2.68	3.05	3.04	3.17	3.07
FRA	2.63	3.16	2.83	3.17	3.21
GEO	2.49	2.86	2.70	3.28	3.66
GRC	2.67	3.73	2.64	3.19	2.78
IRL	2.70	3.04	2.93	3.28	3.04
LUX	2.46	2.74	2.88	3.41	3.51
MLT	2.45	3.45	2.69	3.47	2.94
MKD	2.11	3.30	2.79	3.30	3.50
PRT	2.10	3.44	2.51	3.60	3.36
SRB	2.36	4.02	2.55	3.26	2.82
SVN	2.43	3.51	2.62	3.41	3.04
ESP	2.48	3.25	2.86	3.26	3.15
TUR	2.45	3.88	2.75	3.00	2.92

A Local/regional (subnational) history

B National history

C Regional (supranational) history

D European history

E World history

Table 3.6 – Historical periods covered in history teaching, as indicated by TES respondents (%)[1]

Member state	A	B	C	D	E	F
ALB	46.20	46.28	49.11	41.28	40.71	44.83
AND	33.33	25.00	25.00	33.33	41.67	33.33
ARM	42.57	37.16	48.65	40.09	51.13	27.25
CYP	25.00	25.99	23.03	18.75	35.53	28.95
FRA	29.17	52.27	51.14	53.03	61.36	70.45
GEO	46.21	48.74	49.46	37.18	41.52	34.66
GRC	21.13	32.65	32.47	29.73	50.86	35.40
IRL	34.36	32.52	54.60	57.67	63.19	28.83
LUX	29.25	38.68	39.62	48.11	54.72	63.21
MLT	41.43	34.29	35.71	48.57	35.71	21.43
MKD	60.68	56.53	62.52	49.31	49.16	49.31
PRT	48.11	63.21	67.45	60.85	67.45	67.92
SRB	64.96	67.84	73.42	71.56	72.30	72.96
SVN	76.07	75.21	75.21	72.65	73.50	70.09
ESP	40.98	43.44	41.39	45.08	53.28	47.54
TUR	26.15	22.60	32.33	25.76	30.49	24.84

A Prehistory

B Antiquity

C Middle Ages

D Early modern history

E Modern history

F Contemporary history

Table 3.7 – Frequency of use of methods for teaching and learning history, as indicated by TES respondents, ranging from 1 (never) to 5 (very often)

Member state	A	B	C	D	E	F	G	H
ALB	3.75	3.92	3.58	3.68	3.75	2.95	3.70	3.45
AND	3.17	3.17	3.33	3.67	3.67	3.33	3.76	3.50
ARM	3.94	4.10	3.67	3.66	2.95	3.31	3.70	3.60
CYP	3.48	3.13	2.90	3.06	2.69	2.16	2.97	3.26
FRA	2.39	2.56	2.95	3.07	3.01	2.59	3.50	3.66
GEO	4.15	4.08	3.20	3.38	3.47	3.15	4.09	4.04
GRC	3.79	3.38	3.00	3.14	2.85	2.80	3.34	3.24
IRL	3.75	3.39	3.02	3.13	3.22	2.46	3.64	3.81
LUX	3.51	3.17	3.03	3.19	2.70	2.39	3.29	3.57

1. These values represent the average of binary (0 or 1) variables, so they can be interpreted as the proportion of respondents in each country who selected each field.

Member state	A	B	C	D	E	F	G	H
MLT	3.63	3.39	2.94	3.22	2.90	2.73	3.47	3.69
MKD	4.26	3.72	3.60	3.71	3.58	3.08	3.61	3.33
PRT	2.97	3.36	3.38	3.36	3.12	2.97	3.51	3.76
SRB	4.23	3.48	3.45	3.49	3.07	2.78	3.77	3.39
SVN	3.97	3.62	3.48	3.30	2.88	2.69	3.59	3.46
ESP	4.03	3.39	3.15	3.28	2.99	2.65	3.75	3.54
TUR	4.39	3.56	3.19	3.06	3.00	2.65	3.43	3.07

A Delivering lectures/presentations

B Debating on controversial historical issues

C Questioning how history is represented in the public space (movies, street names, monuments, games, graphic novels, etc.)

D Reflecting on how history is written and used

E Project-based learning

F Place-based learning (outside the classroom, such as visits to museums, historical sites and archives)

G Working with periodisations and timelines

H Using contrasting historical sources and multiple narratives about past events

Table 3.8 – Factors most influential in teaching practice, as indicated by TES respondents, ranging from 1 (least influential) to 5 (most influential)

Member state	A	B	C	D	E
ALB	4.40	3.55	4.0	3.78	3.12
AND	3.00	3.17	2.83	2.83	2.33
ARM	4.14	3.69	4.39	4.01	3.11
CYP	4.00	3.91	2.98	2.72	2.52
FRA	2.74	3.77	2.86	2.78	2.88
GEO	3.55	3.00	3.30	3.04	3.13
GRC	4.36	3.98	2.39	2.24	2.16
IRL	3.23	4.12	3.07	2.63	4.15
LUX	3.55	3.83	2.87	2.87	3.58
MLT	2.65	3.85	2.93	2.85	4.11
MKD	4.19	3.35	3.84	3.55	3.09
PRT	4.05	3.08	3.63	3.41	2.83
SRB	3.98	3.20	3.27	3.00	2.73
SVN	3.87	3.40	3.43	3.60	2.97
ESP	3.06	3.19	3.10	3.08	3.12
TUR	4.25	4.12	3.31	3.20	2.89

A History textbooks

B Exams

C In-service professional development

D Initial teacher training

E Student needs and interests

Table 3.9 – Obstacles to good-quality history teaching, as indicated by TES respondents (%)[2]

	ALB	AND	ARM	CYP	FRA	GEO	GRC	IRL
Time allocated in the curriculum to history	21.16	16.67	45.27	59.21	54.55	25.63	58.08	57.06
Time available to prepare for lessons	8.16	8.33	15.32	32.57	22.73	18.41	27.66	47.86
Availability of qualified teachers	5.90	0.00	15.09	19.74	5.30	5.05	24.40	8.59
Status of history in school	8.64	0.00	17.57	14.80	7.58	3.97	29.04	15.34
Size of the class	10.66	0.00	15.32	39.80	53.79	21.66	37.11	31.29
Resources and budget	29.32	0.00	31.98	19.08	28.79	41.88	28.18	23.31
Curriculum overload	15.67	33.33	38.29	52.96	57.95	32.13	52.41	50.31
Frequency of educational reforms	24.72	41.67	19.14	22.70	38.64	37.91	24.57	20.86
Lack of opportunities for continued professional development	10.10	0.00	15.09	12.50	15.15	14.80	12.71	16.56
Lack of awareness of good practice	3.55	0.00	12.61	23.03	6.82	25.63	25.77	7.98
Focus on the demands of exams and assessment	5.25	8.33	12.39	41.45	37.88	22.02	48.63	43.56

	LUX	MLT	MKD	PRT	SRB	SVN	ESP	TUR
Time allocated in the curriculum to history	58.49	64.29	24.73	69.81	53.16	53.85	43.85	33.38
Time available to prepare for lessons	33.96	35.71	10.91	44.34	10.97	26.50	45.90	12.22
Availability of qualified teachers	16.98	10.00	7.68	7.55	3.16	5.13	6.15	14.32
Status of history in school	33.96	41.43	16.74	34.91	26.02	23.08	15.16	13.01
Size of the class	29.25	34.29	15.67	36.79	38.66	41.88	35.25	11.43
Resources and budget	2.83	24.29	31.18	16.04	25.56	14.53	24.18	17.74
Curriculum overload	44.34	44.29	20.58	31.13	43.96	54.70	40.98	27.20
Frequency of educational reforms	16.04	27.14	35.79	29.25	32.62	4.27	51.23	15.64
Lack of opportunities for continued professional development	4.72	15.71	21.97	16.51	7.43	16.24	11.07	12.22
Lack of awareness of good practice	12.26	7.14	14.13	4.25	9.48	10.26	12.70	9.59
Focus on the demands of exams and assessment	29.25	28.57	7.83	37.26	32.99	31.62	20.49	22.34

2. Note that these values represent the average of binary (0 or 1) variables, so they can be interpreted as the proportion of respondents in each country who indicated each concern.

Table 3.10 – Teachers' views on the relevance of learning outcomes, as indicated by TES respondents, ranging from 1 (least relevant) to 5 (most relevant)

Member state	A	B	C	D	E	F	G	H	I	J	K	L	M	N
ALB	4.1	4.51	4.51	4.5	4.47	4.6	4.53	4.21	4.33	4.26	4.17	4.38	4.34	4.59
AND	3.0	4.17	2.33	3.67	4.0	4.5	4.5	4.33	3.33	3.67	4.0	3.67	3.83	4.5
ARM	3.96	4.68	4.68	3.42	4.49	4.73	4.75	4.21	4.54	4.35	4.43	4.13	4.3	4.42
CYP	2.9	4.66	3.3	3.78	4.7	4.37	4.71	4.29	4.3	4.3	4.55	4.26	4.24	4.66
FRA	3.65	4.11	1.91	2.98	4.53	4.52	4.22	3.58	3.96	4.43	4.16	4.05	3.64	4.48
GEO	3.54	4.63	4.46	4.41	4.78	4.78	4.67	4.33	4.45	4.38	4.57	4.47	4.31	4.66
GRC	2.86	4.62	2.85	3.63	4.6	4.73	4.71	4.19	4.26	4.06	4.5	4.3	4.14	4.66
IRL	3.13	4.61	3.08	3.16	4.5	4.68	4.51	3.96	4.32	4.48	4.66	4.17	4.18	3.89
LUX	3.34	4.39	1.76	3.32	4.57	4.7	4.43	3.71	3.7	4.29	4.38	3.68	4.07	4.38
MLT	2.91	4.48	3.7	3.74	4.61	4.46	4.59	4.11	4.2	4.11	4.39	3.8	4.11	4.0
MKD	3.84	4.53	4.39	4.08	4.48	4.63	4.39	4.02	4.26	4.05	4.23	4.32	4.26	4.32
PRT	3.17	4.64	2.43	3.79	4.65	4.62	4.74	4.45	3.85	4.52	4.71	4.46	4.4	4.77
SRB	3.61	4.6	4.07	3.95	4.56	4.79	4.71	4.27	4.43	4.17	4.48	4.27	4.3	4.46
SVN	3.09	4.36	4.03	3.86	4.47	4.6	4.5	4.11	3.99	4.09	4.36	4.07	4.17	4.54
ESP	2.71	3.76	2.91	2.73	3.67	4.12	3.98	3.46	3.22	3.73	3.8	3.51	3.58	3.83
TUR	4.37	4.54	4.19	3.16	4.37	4.53	4.59	4.3	4.36	4.21	4.4	4.31	3.87	4.33

- A To learn and remember historical facts, dates and processes
- B To recognise and discuss the historical significance/relevance of events and processes
- C To develop national pride
- D To develop a sense of shared European identity
- E To critically analyse historical sources
- F To identify the causes and consequences of historical events and processes
- G To understand and recognise continuity and change in history
- H To understand and reflect on the ethical dimension of history
- I To ask and answer historical questions
- J To contextualise historical events and developments
- K To be aware that there are multiple perspectives in history
- L To learn about multiple identities and cultures that coexisted in the past
- M To learn about historical injustices, including forms of political, social and economic violence against minorities
- N To develop competences for democratic culture

Table 3.11 – Frequency of use of different assessment methods, as indicated by TES respondents, ranging from 1 (never) to 5 (regularly)

Member state	A	B	C	D	E	F	G	H	I	J
ALB	4.05	4.3	3.94	4.01	3.56	3.88	4.21	4.65	3.87	4.08
AND	4.0	3.12	4.0	3.62	2.88	3.88	3.5	3.5	3.62	3.25
ARM	3.3	4.11	3.89	3.94	3.45	3.52	2.7	4.61	3.77	3.76
CYP	3.76	3.6	3.69	2.74	2.72	2.74	3.18	3.94	3.2	2.92
FRA	3.96	3.42	4.27	3.64	2.21	3.16	3.85	3.38	2.98	3.16
GEO	4.36	4.29	4.38	4.15	3.63	3.98	3.83	4.07	3.69	4.05
GRC	3.52	3.56	3.51	2.83	2.79	2.85	3.14	3.76	3.16	2.98
IRL	4.38	4.23	3.85	3.51	2.68	3.05	3.11	4.04	3.55	2.63
LUX	4.33	3.31	3.68	3.57	2.2	2.88	3.48	2.96	3.29	3.28
MLT	4.2	3.8	3.46	3.3	2.83	2.96	3.35	2.74	3.2	2.67
MKD	3.83	4.04	3.76	3.88	3.29	3.82	3.95	4.53	3.67	3.76
PRT	4.47	3.83	4.06	3.8	2.98	3.63	4.01	3.9	3.79	3.75
SRB	3.65	4.18	3.53	3.46	2.83	3.61	3.53	4.58	3.47	3.50
SVN	3.66	3.48	3.44	2.9	2.6	3.23	3.17	4.01	3.1	3.36
ESP	3.79	3.83	3.42	3.7	3.08	3.78	3.32	3.84	3.55	2.71
TUR	3.75	3.93	3.58	3.6	3.32	3.66	3.9	3.71	3.60	3.35

- **A** Exercises that require the interpretation of written and visual historical sources
- **B** Factual questions about historical events or personalities (true/false, multiple choice, linking of dates with events, etc.)
- **C** Essay questions that require argumentation (e.g., causes/consequences, change/continuity, historical interpretations)
- **D** Research tasks where students collect and process information themselves
- **E** Activities (e.g., role play and simulations) where students demonstrate historical empathy
- **F** Project work (e.g., presentations, tours, exhibitions and documentaries)
- **G** Exercises meant to demonstrate understanding of substantive historical concepts (e.g., the Industrial Revolution, modernisation, migration)
- **H** Oral assessment
- **I** Activities that assess student understanding of multiple perspectives on history
- **J** Activities that assess student competences for democratic culture

Sales agents for publications of the Council of Europe
Agents de vente des publications du Conseil de l'Europe

BELGIUM/BELGIQUE
La Librairie Européenne -
The European Bookshop
Rue de l'Orme, 1
BE-1040 BRUXELLES
Tel.: + 32 (0)2 231 04 35
Fax: + 32 (0)2 735 08 60
E-mail: info@libeurop.eu
http://www.libeurop.be

Jean De Lannoy/DL Services
c/o Michot Warehouses
Bergense steenweg 77
Chaussée de Mons
BE-1600 SINT PIETERS LEEUW
Fax: + 32 (0)2 706 52 27
E-mail: jean.de.lannoy@dl-servi.com
http://www.jean-de-lannoy.be

CANADA
Renouf Publishing Co. Ltd.
22-1010 Polytek Street
CDN-OTTAWA, ONT K1J 9J1
Tel.: + 1 613 745 2665
Fax: + 1 613 745 7660
Toll-Free Tel.: (866) 767-6766
E-mail: order.dept@renoufbooks.com
http://www.renoufbooks.com

FRANCE
Please contact directly/
Merci de contacter directement
Council of Europe Publishing
Éditions du Conseil de l'Europe
F-67075 STRASBOURG Cedex
Tel.: + 33 (0)3 88 41 25 81
E-mail: publishing@coe.int
http://book.coe.int

Librairie Kléber
1, rue des Francs-Bourgeois
F-67000 STRASBOURG
Tel.: + 33 (0)3 88 15 78 88
Fax: + 33 (0)3 88 15 78 80
E-mail: librairie-kleber@coe.int
http://www.librairie-kleber.com

NORWAY/NORVÈGE
Akademika
Postboks 84 Blindern
NO-0314 OSLO
Tel.: + 47 2 218 8100
Fax: + 47 2 218 8103
E-mail: support@akademika.no
http://www.akademika.no

POLAND/POLOGNE
Ars Polona JSC
25 Obroncow Street
PL-03-933 WARSZAWA
Tel.: + 48 (0)22 509 86 00
Fax: + 48 (0)22 509 86 10
E-mail: arspolona@arspolona.com.pl
http://www.arspolona.com.pl

PORTUGAL
Marka Lda
Rua dos Correeiros 61-3
PT-1100-162 LISBOA
Tel: 351 21 3224040
Fax: 351 21 3224044
E-mail: apoio.clientes@marka.pt
www.marka.pt

SWITZERLAND/SUISSE
Planetis Sàrl
16, chemin des Pins
CH-1273 ARZIER
Tel.: + 41 22 366 51 77
Fax: + 41 22 366 51 78
E-mail: info@planetis.ch

UNITED KINGDOM/ROYAUME-UNI
Williams Lea TSO
18 Central Avenue
St Andrews Business Park
Norwich
NR7 0HR
United Kingdom
Tel. +44 (0)333 202 5070
E-mail: customer.services@tso.co.uk
http://www.tsoshop.co.uk

UNITED STATES and CANADA/
ÉTATS-UNIS et CANADA
Manhattan Publishing Co
670 White Plains Road
USA-10583 SCARSDALE, NY
Tel: + 1 914 472 4650
Fax: + 1 914 472 4316
E-mail: coe@manhattanpublishing.com
http://www.manhattanpublishing.com

Council of Europe Publishing/Éditions du Conseil de l'Europe
F-67075 STRASBOURG Cedex
Tel.: + 33 (0)3 88 41 25 81 – E-mail: publishing@coe.int – Website: http://book.coe.int

Why do we learn history?

History education is increasingly recognised for its contribution to democratic citizenship education. Knowledge of the past is important not only for its own sake but also for developing young people's analytical and critical thinking skills. It should not only provide them with factual information, but it should also develop their historical thinking, thus allowing them to become informed, active citizens.

When taught according to quality history teaching precepts, history as a subject matter can play a crucial role in building and maintaining democratic societies.

How history is taught matters.

The mission of the Observatory on History Teaching in Europe (OHTE) is to provide a clear picture of the state of history teaching in its member states. This is done through OHTE general reports on the state of history teaching in Europe and OHTE thematic reports, which explore particular areas of interest and how they are treated in history lessons.

This first OHTE *General report on the state of history teaching in Europe* captures the state of history teaching in the OHTE member states. It covers topics such as the place of history in education systems, thematic foci within curricula, the use of history textbooks and other educational resources, preferred pedagogical practices, learning outcomes and assessment, as well as information on history teachers and their training.

www.coe.int/ohte
X @CoE_History
Facebook @OHTECOE

PREMS 002624

ENG

www.coe.int

The Council of Europe is the continent's leading human rights organisation. It comprises 46 member states, including all members of the European Union. All Council of Europe member states have signed up to the European Convention on Human Rights, a treaty designed to protect human rights, democracy and the rule of law. The European Court of Human Rights oversees the implementation of the Convention in the member states.

OBSERVATORY ON HISTORY TEACHING IN EUROPE

COUNCIL OF EUROPE
75
1949 - 2024
CONSEIL DE L'EUROPE