



BMJ Open A cross-sectional study on adaptation and initial validation of a test to evaluate health claims among high school students: Croatian version

Diana Aranza,¹ Boris Milavić,² Ana Marusic ,^{3,4} Magdalena Buzov,⁵ Tina Poklepović Peričić ^{3,4}

To cite: Aranza D, Milavić B, Marusic A, *et al.* A cross-sectional study on adaptation and initial validation of a test to evaluate health claims among high school students: Croatian version. *BMJ Open* 2021;**11**:e048754. doi:10.1136/bmjopen-2021-048754

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-048754>).

Received 27 February 2021
Accepted 28 June 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹University Department for Health Studies, University of Split, Split, Croatia

²Department of Anthropological Kinesiology and Health, University of Split Faculty of Kinesiology, Split, Croatia

³Research in Biomedicine and Health, University of Split School of Medicine, Split, Croatia

⁴Cochrane Croatia, University of Split School of Medicine, Split, Croatia

⁵Study of Dental Medicine, University of Split School of Medicine, Split, Croatia

Correspondence to

Dr Tina Poklepović Peričić; tpoklepo@mefst.hr

ABSTRACT

Objectives We validated the Croatian version of the test using multiple-choice questions (MCQs) from the Claim Evaluation Tools item bank of the Informed Health Choices project, and measured the ability of high school students to appraise health claims.

Setting 16 high schools from the urban agglomeration of the city of Split, Croatia.

Participants Final year high school students of at least 18 years of age.

Interventions 18 MCQs from the item bank considered relevant for high school students were translated. After face-validity testing, the questionnaire was piloted and sent to a convenient sample of 302 high school students.

Primary and secondary outcome measures Difficulty and discrimination indices were calculated for each MCQ to determine the validity of translation and the weight of MCQs. We assessed basic metric characteristics and performed initial validation of the test. Two tests were created, the full (18 MCQs) and the short version (12 MCQs). We analysed differences in test score according to gender and school.

Results The response rate was 96% (75% female respondents). Metric characteristics of both tests were satisfactory (Cronbach's $\alpha=0.71$ for the full and $\alpha=0.73$ for the short version). The mean score (\pm SD) for the full version was 11.15 ± 3.43 and 8.13 ± 2.76 for the short version. There were 6 easy and 12 moderately difficult questions. Questions concerning effectiveness and dissimilar comparison groups were answered correctly by fewer than 40% of students. Female students and those from grammar and health schools scored higher on both tests.

Conclusions Both tests showed good metric characteristics and may be used for quick and reliable assessments of adolescents' ability to appraise health claims. They may be used to identify needs and inform development of educational activities to foster critical thinking about health among adolescents.

BACKGROUND

The average time children spend using different kinds of media is increasingly growing over time.¹ Whether through the internet, social media, television or

Strengths and limitations of this study

- This is the first initiative to validate a test for measuring ability of high school students to assess claims about treatment effects in the Croatian language.
- The study included a diverse sample of final year high school students from 16 schools with various teaching programmes.
- The test was developed using the items from the Claim Evaluation Tools item bank of the Informed Health Choices project, followed by feedback from methodologists and teachers, and a piloting exercise involving high school students.
- We determined the quality, understandability and the weight of the test questions using difficulty index and discrimination index, assessed basic metric characteristics and performed initial validation of the Croatian version of the test including assessment of homogeneity, reliability, sensitivity and validity.
- Limitations of this study include predominance of female student respondents, regional sample and the lack of controlling for other variables that may influence assessing health claims.

magazines, children, like adults are exposed to various claims about the benefits and harms of treatments, which implies an urgent need to empower critical thinking about health among children and adolescents.²⁻⁶ Health claims shared through media are most often related to quick and easy solutions for health problems or their prevention, where most of them are unreliable and carry a risk of human suffering and unnecessary costs.² Health literacy is defined as 'the degree to which individuals have the capacity to obtain, process and understand basic health information needed to make appropriate healthcare decisions'.⁷ Low levels of health literacy have been shown to contribute to health inequalities and disparities,⁸⁻¹⁰ and are considered a predictor of health status stronger than age, education, employment status, income and

ethnicity.¹¹ The Informed Health Choices (IHC) project is an international, multidisciplinary team of experts who have developed a list of 49 key concepts as a framework for thinking about health claims in everyday decision making.¹² The IHC project also developed educational materials for teaching in primary school settings, along with assessment tools.¹² The 49 key concepts are organised within the following three main groups of concepts: (i) claims, referring to recognising claims about treatment effects that have unreliable basis, (ii) comparisons, emphasising the importance of fair comparisons of treatments and (iii) choices, referring to the process of judging the relevance of the evidence, balancing benefits, harms and cost when making health decisions.^{13 14} The IHC project points out that good health and quality of life mostly depend on people's ability to understand and appraise health information, and make the right choices about their health.¹⁵ The Claim Evaluation Tools item bank, developed within the IHC project is a set of validated multiple-choice questions (MCQs) for assessing the level of understanding and the ability to apply the key concepts in decision making. Besides, the Claim Evaluation Tools item bank may be used to develop tests for schools, or to assess the effects of educational interventions. The items available from the item bank encompass a wide range of topics and difficulty levels to investigate the ability to critically appraise health claims in various population groups. This may allow insight in the status of the population and development of specific educational interventions that would satisfy people's needs.¹⁵

The National Program for Health Education in Croatia, introduced in 2014, emphasises the need for a critical approach in making decisions about health. According to the national curriculum, the aim of health education is to enable students to make responsible decisions for nurturing their physical and mental health, to question their own behaviour and make sound health choices.¹⁶

The main objective of our study was to adapt and validate a test made of a set of MCQs from the Claim Evaluation Tools item bank, translated into Croatian. The initial validation aimed to establish basic metric characteristics of the test (homogeneity, confidence, sensitivity and validity) for measuring the ability of high school students to appraise health claims, as well as to create a shorter, economic version of the test, without losing the quality of metric characteristics of the full test. Additionally, we aimed to determine the level of critical appraisal skills among high school students, and analyse possible determinants of critical thinking, which may provide basis for planning specific educational and preventive programmes aimed to foster critical thinking among adolescents.

METHODS

Study design

We performed a cross-sectional, questionnaire-based study that involved a convenient sample of final year high

school students from the urban agglomeration of the city of Split in the south of Croatia, with the population of a quarter of a million. The study was conducted during May and June 2020 and was reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting checklist.¹⁷

Development of the test

From the full set of questions available in the Claim Evaluation Tools item bank, two authors (DA, TPP) independently selected items referring to the Key Concepts that they considered eligible for high school students in the current health education model. The two authors compared their lists of selected questions, discussed discrepancies and agreed on a set of 18 items to include in the questionnaire. All questions were translated into Croatian and back-translated into English to check content validity of the translation. Another author (AM), not involved in selecting the questions, assessed the test and provided feedback on relevancy, understandability and appropriateness of the selected questions. We sent the test to five other experts, methodologists and physicians with experience in research, for independent feedback. We asked the experts to complete the test, without having revealed the correct answers to them. Also, we asked them to provide comments on any problem they might encounter while answering questions. If for any reason, one or more methodologist found any of the questions doubtful, we revised the questions for clarity. All experts completed the test correctly. Based on their feedback, we made changes in wording to make more sense of the questions and to put the questions in the local context. The first MCQ related to the Key Concept '100% safe!' was removed and replaced with another question from the Claim Evaluation Tools item bank from the same group of questions, because two experts had comments regarding the ambiguity of the answers. After addressing the experts' feedback, we sent the 18 MCQs for feedback to four high school teachers from one grammar school, the health school, the school for graphic and design and the economy high school. Besides minor comments on terminology, which we addressed, all four teachers found the questions to be clear and easy to understand.

Setting and sampling

Before the start of the study, we sent formal emails to the principals of 20 high schools with whom we managed to establish contacts, considering that schools were closed and switched to online teaching due to COVID-19 pandemic. We informed them about the aims and design of the study, and invited them to take a part in the study. Principals of the 16 high schools expressed interest in the study and were informed about the details of the study. We contacted teachers from the 16 high schools whose principals agreed to participate in this study via email. The target population were students of at least 18 years of age, that is, adults who can provide informed consent.

Power analysis

The number of students was based on the number of participants used in previous similar studies, in which a sample of 300 participants was considered appropriate to test a questionnaire of approximately 24 questions.^{15 18} Having considered the number of final year students in the participating schools, we agreed that at least one class of students per school would be sufficient to reach a satisfactory sample size.

Data collection

We developed an online questionnaire using a Google Docs form. The form contained information about the scope of the study that students had to go through before answering questions. Completing the questionnaire was considered as consent for participation. The first part of the questionnaire included demographic data, like age, gender and the type of school participants attended. Each of the 18 MCQs consisted of a scenario leading to a treatment claim and a question with 3 or 4 possible answers.

First, we piloted the questionnaire on a sample of 15 grammar school final year students. We collected data regarding the frequency of correct answers and the time needed to complete the test, and asked them for their feedback. All students reported having no problem with completing the test, found it clear and easy to complete. The average time needed to complete the test was around 20 minutes. Data obtained from the pilot test were not included in the analysis. Teachers who agreed to participate were sent the link to the online questionnaire which they forwarded to their students. All answers from the test were automatically imported in an Excel file (Ver. Office 2007, Microsoft, Redmond, Washington, USA). We coded the answers and analysed the data using the SPSS V.24 (IBM, Armonk, New York, USA).

Validation of the test

The quality of MCQs in English from the Claim Evaluation Tools item bank was tested using the Rasch Analysis modelling.¹⁸ We determined the quality, understandability and the weight of the translated questions using difficulty index and discrimination index.¹⁹ We also assessed basic metric characteristics and performed initial validation of the Croatian version of the test: homogeneity—establishment of the latent structure of the test was done by applying the exploratory principal component analysis (PCA) with Varimax orthogonal rotation and Guttman-Kaiser criteria for determining the number of significant components; reliability—by calculating the internal consistency coefficient of the Cronbach's α type; sensitivity—by calculating the Kolmogorov-Smirnoff (K-S) goodness-of-fit coefficient to test for the normality of distribution, along with other sensitivity indices, like the range of results through minimal and maximal result; skewness and kurtosis coefficients; validity—by assessing discriminative validity through differentiating between subsamples. Discriminative indices of the questions were calculated by comparing the proportions of correct

answers with questions between subgroups of participants who had the highest or the lowest overall test score.^{20 21} To develop a short version of the test, we used a process of item selection procedure, by which we eliminated those questions that had overlapping content, or were low in contributing to some of the metric characteristics (homogeneity of the latent structure of the test, reliability or test sensitivity).

Data analysis

We used descriptive statistics to present data from the test regarding the demographics, as well as frequencies of correct answers, and have presented those as absolute numbers and percentages for each test question. The overall test scores were presented using means (M) and SD. For each question, we assessed gender differences using the Mann-Whitney U test. Gender differences in the overall test results were analysed using the t-test for independent samples. The variance of the overall test results regarding the type of school was assessed using the one-way analysis of variance test. P value <0.05 was considered statistically significant. Fisher's least significant difference (LSD) test was used for post hoc analyses of the differences in the overall test scores between groups of female and male students in relation to the type of school.

Multiple regression analyses were performed to investigate the overall multivariate effect of the two variables, gender and the type of school, as well as to determine the level of individual contribution of each of these variables on the overall test results.

The items from the Claim Evaluation Tools item bank are not publicly available,²² but access can be obtained from the IHC project leaders on request. The online supplemental file contains the list of Key Concepts of the questions selected for the Croatian test.

Patient and public involvement

During the development of the test, medical experts and high school teachers helped in assessing the appropriateness of the selected questions, and provided feedback. A sample of 15 high school students were involved in piloting the test. The participants in the study were high school students, and the study was designed by a team of researchers from the University of Split.

RESULTS

Participants

From the total of 302 high school students included in this study, 13 students failed to provide answers to all questions and were excluded from the analyses, leaving 289 tests (95.7% response rate) for analyses. There were 216 (75%) female and 73 (25%) male students, with a mean age of 18.3±0.6; 101 students attended grammar schools, 98 health high school and 90 vocational high schools (economy, graphics, catering, maritime, construction and geodesy, and high schools with mixed programmes from two islands).

Table 1 Item difficulty, discrimination indices and item—test score correlation of the Croatian version of the Claim Evaluation test

Multiple-choice questions (MCQs)	Difficulty index			Discriminative index				Correlation with the overall score*
	Incorrect answer (IA)	Correct answer (CA)	Index	Highest results group (n=86)	Lowest results group (n=76)	High-low difference	Index	
	N (%)	N (%)		CA proportion	CA proportion			
MCQ_1	51 (17.7)	238 (82.4)	Easy	0.95	0.63	0.32	Good	0.23†
MCQ_2	185 (64.0)	104 (36.0)	Moderately	0.63	0.22	0.43	Very good	0.19†
MCQ_3	141 (48.8)	148 (51.2)	Moderately	0.70	0.36	0.34	Very good	0.10
MCQ_4	106 (36.7)	183 (63.3)	Moderately	0.96	0.24	0.72	Very good	0.46†
MCQ_5	98 (34.0)	190 (66.0)	Moderately	0.84	0.41	0.43	Very good	0.29†
MCQ_6	133 (46.0)	156 (54.0)	Moderately	0.86	0.29	0.57	Very good	0.33†
MCQ_7	110 (38.1)	179 (62.0)	Moderately	0.91	0.43	0.48	Very good	0.26†
MCQ_8	166 (57.4)	123 (42.6)	Moderately	0.66	0.29	0.37	Good	0.14‡
MCQ_9	104 (36.0)	185 (64.0)	Moderately	0.86	0.35	0.51	Very good	0.29§
MCQ_10	83 (28.7)	206 (71.3)	Easy	0.91	0.42	0.49	Very good	0.34†
MCQ_11	79 (27.3)	210 (72.7)	Easy	0.97	0.35	0.62	Very good	0.52†
MCQ_12	53 (18.3)	236 (81.7)	Easy	0.96	0.58	0.38	Good	0.38†
MCQ_13	77 (26.6)	212 (73.4)	Easy	0.97	0.40	0.57	Very good	0.46†
MCQ_14	183 (63.3)	106 (36.7)	Moderately	0.53	0.22	0.31	Good	0.14‡
MCQ_15	104 (36.0)	185 (64.0)	Moderately	0.88	0.38	0.50	Very good	0.29†
MCQ_16	119 (41.2)	170 (58.8)	Moderately	0.87	0.30	0.57	Very good	0.37†
MCQ_17	106 (36.7)	183 (63.3)	Moderately	0.78	0.44	0.34	Good	0.18‡
MCQ_18	82 (28.4)	207 (71.6)	Easy	0.88	0.49	0.39	Good	0.25†

*Pearson's coefficient of correlation between results for individual questions and overall test score.

†Statistically significant coefficient at the level of $p < 0.001$.

‡Statistically significant coefficient at the level of $p < 0.01$.

§Statistically significant coefficient at the level of $p < 0.05$.

Frequency of correct answers and association with the overall test score

Table 1 presents the frequencies of correct answers for each question, item difficulty index and discrimination index, as well as the correlation of each question data with the overall test score.

Percentages of correct answers to specific test questions varied from 36.0% to 82.4%. Difficulty index of the questions was lowest for the 1st and the 12th question, with both having over 80% of correct answers. Overall, there were six easy questions, while all other questions were considered as moderately difficult. Discriminative index of all questions varied from good to very good. The results of all questions were significantly associated with the overall test result except for the third question. The Pearson's type coefficients of correlation for questions with the overall score varied between 0.14 (questions 8 and 14) and 0.52 (question number 11).

Factor analysis

Based on the data from the full test, we conducted factor analysis to investigate the main components of the test.

The questions projected to six latent dimensions, with 51.4% of the common variance. Additional item selection was undertaken to minimise the number of questions in the test to allow time saving, and to simplify the latent structure of the test, but without losing much of the metric quality compared with the full test. Item selection resulted in the removal of 6 questions, leaving 12 MCQs for the shorter test version (table 2).

PCA for the short test version showed that the latent structure consisted of three components, explaining from 17.5% to 13.4% of the common variance. Overall, these three components explained as much as 44.8% of the overall variance. The first factor accounted for the total of four questions, with its basis in questions 10 and 12. The second factor encompassed four questions, with the basis in questions 6 and 9. The third factor, with three questions, had a basis in questions 1 and 3. Correlations of each specific question from the short test version with the overall score were significant, of which four were moderate (above 0.40).

Table 2 Factor structure of the short version of the Croatian CLAIMS test

Multiple-choice questions (MCQs)	Principal component analysis*			Question-total correlation*
	Factor 1	Factor 2	Factor 3	
MCQ_1	0.03	-0.18	0.78	0.21
MCQ_4	0.18	0.25	0.64	0.42
MCQ_5	0.36	0.16	0.25	0.32
MCQ_6	0.12	0.74	-0.06	0.33
MCQ_7	0.11	0.16	0.48	0.28
MCQ_9	-0.01	0.66	0.19	0.31
MCQ_10	0.68	0.09	0.03	0.38
MCQ_11	0.57	0.25	0.34	0.54
MCQ_12	0.82	0.05	-0.03	0.42
MCQ_13	0.58	0.15	0.29	0.46
MCQ_15	0.27	0.39	0.01	0.28
MCQ_16	0.19	0.54	0.18	0.37
Explained variance	2.10	1.68	1.60	Total variance explained=44.8%
% of variance	17.5	14.0	13.4	

*Factor saturations of test questions.

Descriptive values of the overall test results

The total mean scores of the test, both for the full and for the short test version were close to the corresponding median values of the full and the short test version (table 3). The mean value was approximately 62% of the maximum score in the full test version and 68% in the short version. The Cronbach's α coefficient of internal consistency for both tests reached a satisfactory level for the full and for the short test version (table 3). In general, sensitivity indicators for the overall results of both test versions were good, although the results of the K-S test indicated significant deviation in the results from a normal distribution. Namely, overall results for both versions of the test were scattered across the whole possible range of results, with measures of skewness and kurtosis for the distribution of results being in a range of ± 1.00 (table 3), suggesting that the metric and descriptive characteristics of both tests were of good quality and that we were allowed further use of parametric tests.

Gender differences in relation to the test results

Female respondents had more correct answers than male respondents, reaching >20% difference of correct

answers to the first question, and >15% difference to male students in answers to the fourth question (table 4). There were statistically significant differences between male and female students regarding questions numbers 1, 3 and 4. The first and the fourth question, with ascertained differences between male and female students, were found to be projected into the third latent component. Likewise, the differences between female and male students were very close to the limit of significance for questions numbers 10, 11 and 12. While analysing the differences in the overall test scores between male and female students for both tests, we also found significant differences. Female students showed a significantly higher level of knowledge measured by both versions of the test than male students. In relation to the full version of the test, the short test showed slightly higher discriminative value.

Female students reached higher scores on both test versions than their male counterparts (figure 1).

In order to explore the potential impact of different educational programmes in high school, we carried out the analysis of variance according to the type of high school (figure 2).

Table 3 Descriptive and basic metric characteristics of the overall result of the test

Variable	M	SD	Median	MIN	MAX	SKEW	KURT	K-S D test	Cronbach's α
Full version (18 MCQs)	11.15	3.43	12.00	2.00	18.00	-0.48	-0.42	0.12*	0.71
Short version (12 MCQs)	8.13	2.76	9.00	1.00	12.00	-0.59	-0.59	0.15*	0.73

*Significant K-S D test.

Cronbach's α , coefficient of internal consistency of the test; K-S D test, Kolmogorov-Smirnov coefficient of deviation from normal distribution; KURT, coefficient of kurtosis of distribution; M, mean; MAX, maximal value; MCQs, multiple-choice questions; MIN, minimal value; SKEW, coefficient of skewness of distribution.

Table 4 Gender differences in the multiple-choice questions (MCQs) and of the overall test score

Multiple choice questions	Female	Male	Mann-Whitney U test		
	%	%	U	Z-adjusted	P value
MCQ_1	87.5	67.1	6277.5*	3.94	0.000
MCQ_2	35.7	37.0	7778.5	-0.21	0.84
MCQ_3	54.6	41.1	6817.0†	2.00	0.046
MCQ_4	67.1	52.1	6695.5†	2.31	0.021
MCQ_5	67.6	60.3	7307.0	1.14	0.26
MCQ_6	55.1	50.7	7536.5	0.65	0.51
MCQ_7	63.9	56.2	7275.0	1.17	0.24
MCQ_8	44.0	38.4	7440.5	0.84	0.40
MCQ_9	66.2	57.5	7200.5	1.33	0.18
MCQ_10	74.1	63.0	7012.0	1.80	0.07
MCQ_11	75.0	65.8	7155.0	1.53	0.13
MCQ_12	83.8	75.3	7217.5	1.61	0.11
MCQ_13	74.1	71.2	7660.0	0.47	0.64
MCQ_14	34.7	42.5	7273.5	-1.18	0.24
MCQ_15	65.3	60.3	7489.5	0.77	0.44
MCQ_16	58.8	58.9	7875.5	-0.02	0.99
MCQ_17	62.5	65.8	7627.5	-0.50	0.62
MCQ_18	72.7	68.5	7553.5	0.69	0.49

Test score	Female		Male		t-test	P value
	M	SD	M	SD		
Full version	11.43	3.33	10.33	3.62	2.38†	0.02
Short version	8.38	2.70	7.38	2.80	2.71*	0.007

*Statistically significant coefficient at the level of $p < 0.01$.

†Statistically significant coefficient at the level of $p < 0.05$.

M, mean; MIN, minimal value; p value, level of significance of the t-test coefficient; SD, standard deviation; t-test, t-test coefficient.

The analysis showed significant differences in test scores regarding the type of school students attended for the full ($F=9.99$; $p < 0.001$) and the short test version ($F=11.67$; $p < 0.001$). Students from the health school reached highest scores (mean \pm SD) at both the full (11.92 \pm 3.07) and the short test (8.89 \pm 2.47). They were followed by grammar school students with mean scores for the full (11.54 \pm 3.41) and the short test (8.36 \pm 2.68). The lowest mean scores for the full (9.87 \pm 3.51) and the short test (7.06 \pm 2.83) were observed among students from vocational schools, who

clearly presented a different population of students, as the CIs for their test results did not overlap with the CIs for students' scores from other two types of schools (figure 2).

We carried out a post hoc analysis using the Fisher's LSD test to explore the level of differences between different types of schools. The students attending other types of schools showed significantly lower scores at both tests compared with students attending health school and grammar schools ($p=0.001$ for the full and $p < 0.001$ for the short test). There was no significant difference in test scores between students from the health school and grammar schools in the short and the full test results ($p=0.16$ and $p=0.43$, respectively).

Results of the multiple regression analyses (table 5) suggested that both variables, gender and type of school correlated with the overall test scores, however both of them were explaining only 6% of the overall variance of the results.

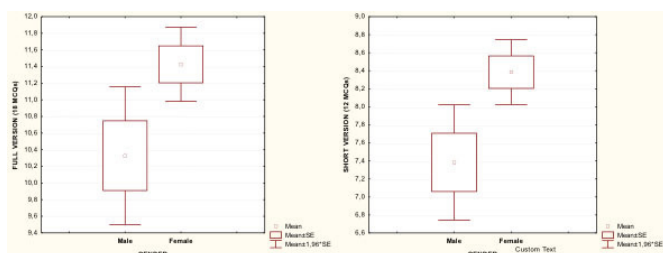


Figure 1 Overall scores on the full and short version of the Croatian version of the Claim Evaluation test according to the students' gender.

DISCUSSION

Study findings

The validation of the Croatian version of the Claim Evaluation test showed initial homogeneity and satisfactory

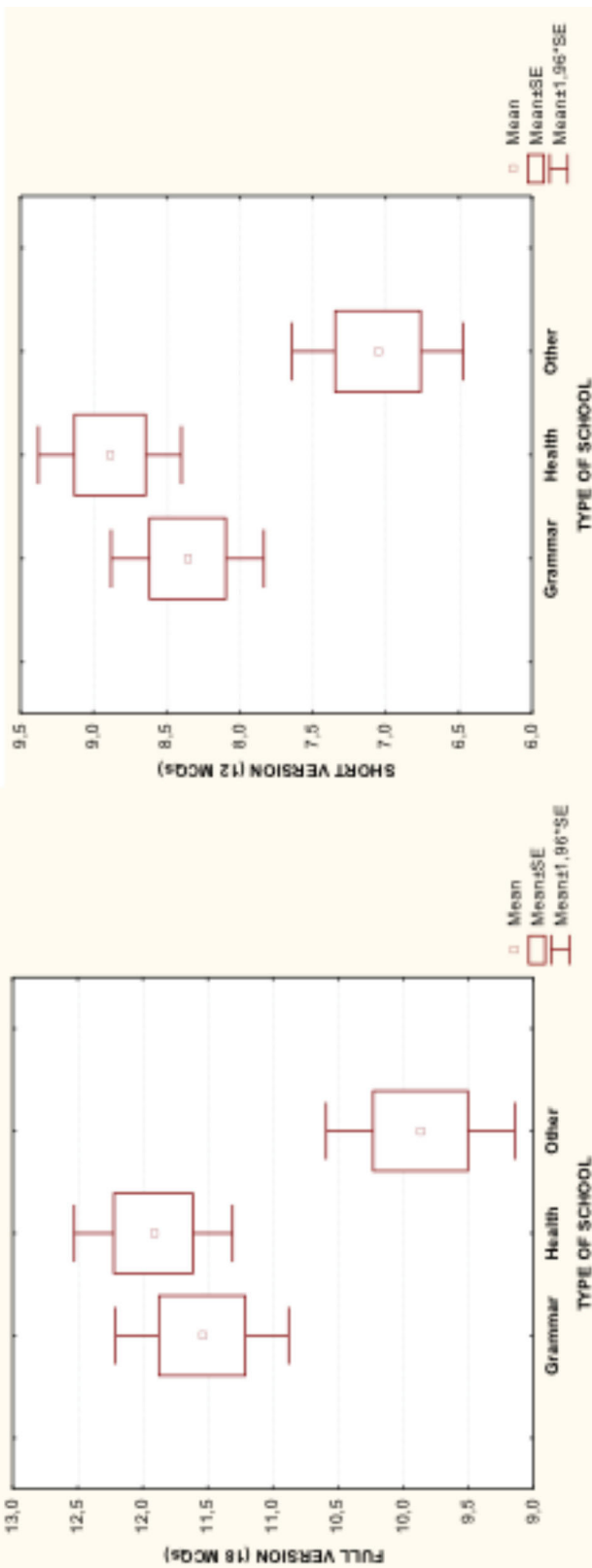


Figure 2 Overall scores on the full and short version of the Croatian version of the Claim Evaluation test according to the high school attended by the study participants.

basic metric characteristics. Both the full and the short test had good discriminative validity for differentiating subgroups of participants according to gender and the

type of school they attended. This implies that both tests can be used for assessing competencies of high school students in critical appraisal of health claims.

Overall, 60% and 70% of the maximal score was achieved in the full and short test, respectively, suggesting that high school students are moderately skilled in critically appraising health claims, with a small number of high school students showing high level of critical appraisal competency.

None of the questions was very difficult for high school students, which means that the questions used in the Croatian test version are of appropriate difficulty for the target population.

The frequency of correct answers varied between questions, and those related to the concepts ‘100% safe!’ and ‘As advertised’ were most often correctly answered, and were therefore considered easiest. On the other hand, questions related to the concepts ‘100% effective!’ and ‘Dissimilar comparison groups’ were most often incorrectly answered, and were therefore considered hardest for high school students. The concept ‘Dissimilar comparison groups’ emphasises the importance of comparing groups in which participants are as similar as possible, otherwise the observed effect may be due to differences between participants, and not to the effect of the investigated treatment.¹⁴ The low rates of correct answers may reflect the lack of training students had in that sense during their education. The concept ‘100% effective!’ describes how rare very effective treatments occur, and that making conclusions about 100% treatment effect is mostly wrong.¹⁴ We noticed that, when answering the question related to the concept ‘100% effective!’, students played safe and more often chose the ‘partially’ incorrect answer than having correctly and completely rationally doubted the reported claim. This, however reveals their uncertainty towards appealing claims, and indicates the need for training on this matter.

Furthermore, we found that both gender and the type of the school were associated with the overall test score. Female students scored higher than male students on both test versions. Students attending the health school scored highest, followed by the students from one grammar school, with students from vocational schools scoring lowest at both tests.

The observed differences in scores regarding different schools may reflect different educational programmes, especially in settings where health topics have higher relevance, like health school. Also, health and grammar schools are likely to have more female than male students. There is evidence that women read blogs about health more often than men.²³ Women are also more likely to leave their jobs to provide care for family members as informal caregivers.^{24 25} Results in favour of female students may also be because they often attend more demanding educational school programmes, and thus reach higher level of critical thinking and knowledge in general. However, the observed gender differences in our study may be due to the overall larger numbers of female students in the study sample.

**Table 5** Multiple regression for both test versions according to gender and type of school

Model	Full version				Short version			
R	0.24				0.24			
R square	0.06				0.06			
Adjusted R square	0.05				0.05			
Std. error of the estimate	3.35				2.68			
F (2286)	8.60*				9.01*			
Sig. F	0.000				0.000			

Variable	Unstandardized coefficients		Standardised coefficients		P level	Unstandardized coefficients		Standardised coefficients	
	B	Std.Err.	Beta	P level		B	Std.Err.	Beta	P level
(Constant)	13.03	0.53	–	0.000	9.62	0.42	–	0.000	
GENDER	–1.10	0.45	–0.14	0.016	–1.00	0.36	–0.16	0.006	
Type of school	–0.82	0.24	–0.19	0.001	–0.63	0.19	–0.19	0.001	

*Statistically significant coefficient at the level of $p < 0.001$

Adjusted R square, adjusted multiple determination coefficient; B, unstandardized regression coefficient; Beta, standardized regression coefficient; F (df 1, df 2), coefficient of analysis of variance (degrees of freedom 1, and 2); p-level, level of significance of the Beta coefficients; R, multiple correlation coefficient; R square, multiple determination coefficient; Sig. F, level of significance of the F coefficient.

Comparison with other studies

A cross-sectional study that surveyed 520 college students in Jordan found that female students showed higher levels of health literacy than their male colleagues.²⁶ Two studies investigated the level of critical appraisal skills about health among Norwegian adults and did not find significant gender differences.^{27 28} In the first study, the average number of correct answers on the test were 4.92 out of the total of 9 questions, showing that there were some important gaps in adults' understanding of the key concepts that are considered necessary for making well-informed decisions about health.²⁷ The second study confirmed insufficient understanding of the key concepts among adults in Norway, with at least half of the adults having understood 18 out of 30 key concepts.²⁸

Several other studies have so far used different items from the Claim Evaluation Tools item bank on different population groups. Validation of the questionnaire in Spanish conducted on adults in Mexico selected 22 of the questions from the Claim Evaluation Tools item bank,²⁹ while the validation in Mandarin used 21 questions.³⁰ The same questions have been used for the studies conducted in Uganda and in Norway.¹⁸

The findings of our study confirm the quality of the Croatian test as a flexible tool for measuring the ability of adolescents to critically appraise health claims. The three latent dimensions that the short test version projected to represent a high-quality basis for measuring critical thinking of high school students. However, they somewhat differ from the structure of the originally proposed groups of concepts, which relate to recognising claims that have an unreliable basis, the importance of fair comparisons and choices in decision making.^{14 31} The internal

reliability of the test, as well as the sensitivity indicators were at a satisfactory level,³² while the deviation from the normal distribution in the results may be explained by high sensitivity of the K-S test to sample size.³³

Strengths and limitations

The strength of this study is that we have developed tests that may be used in a variety of settings to provide quick and reliable assessments of adolescents' ability to appraise treatment claims and make decisions about health. In times of social media and the internet, where adolescents spend a lot of their time, and obtain most of the information, testing their ability to appraise health claims and make decisions about health is important. Final year high school students are at the stage in life that represents a crucial turning point. Some of the high school students will continue their education at the university level, but for some high school represents the final stage of their formal education, so it is important to intensively teach and assess critical thinking about health claims among this population.

We validated the tests using robust methods, but the results should be further confirmed. One of the limitations to our study includes over-representation of female students in the sample, and the lack of controlling for other variables that may influence assessing health claims.

Implications and future research

Both tests from our study can be used as a quick and high-quality assessments tool for health-related critical appraisal skills. Such studies would provide information about the similarities and differences in adolescents' critical appraisal skills across different countries and settings.

The tests could also be used to measure the effects of educational interventions on critical thinking about health that best meet the needs of different student groups.

Furthermore, although the findings of our study suggest that the overall test score may be associated with gender and the type of school, the investigation into other factors contributing to these differences and the overall test score implies that future studies use larger and balanced samples, including variables like overall literacy, socioeconomic status or experience with a health problem, either personally or through a family member. Other sources of knowledge, like family, education, internet and the media, official medical sources, pharmaceutical companies marketing, self-help literature, influencers or celebrities should also be considered in future studies.

The observed good quality metric characteristics of the Croatian test emphasise the appropriateness and good metric characteristics of the questions originally created for the Claim Evaluation Tools item bank 2020.²² In that regard, this study represents continuation of the global endeavours of the international IHC network to create tools for different settings and populations.

CONCLUSIONS

Both the full and the short Croatian version of the Claim Evaluation test are of good quality and allow satisfactory assessment of the level of critical thinking among high school students. The tests can be translated and used worldwide for assessing ability of adolescents to appraise health claims.

High school students in Croatia showed a moderate level of critical thinking towards health claims, with female students and those from health school being more skilled than male students from vocational schools in critical appraisal of health claims. The tests can be used for future research within the global IHC network, and in the high school setting to identify needs and develop action plans carefully tailored to foster critical appraisal of health claims in specific adolescents' groups.

Twitter Ana Marusic @ana_marusic and Tina Poklepović Peričić @Tina Poklepovic Pericic

Acknowledgements We would like to thank Andy Oxman and Astrid Austvoll Dahlgren for their support and kind advice during the planning and conduct of this study. We are thankful to teachers, principals and students from the grammar schools, the health school and the vocational high schools from the Split-Dalmatia County who took part in this study. We thank our colleagues from Cochrane Croatia for their valuable comments on the test: Andrija Babić, Ivan Buljan, Ana Jerončić, Ana Poljičanin, Livia Puljak.

Contributors DA, BM, AM, TPP: planning study design and conduct. DA, MB, TPP: data collection. DA, BM, AM, TPP: data analysis. DA, MB, AM, BM, TPP: writing the manuscript and approving the final version of the manuscript.

Funding The authors received no specific funding for this work. Publishing of the study was funded by the Croatian Science Foundation Project 'Professionalism in Health: Decision making in practice and Science (ProDem)' (IP-2019-04-4882) and the institutional project 'Promoting health literacy in children and adolescents' (SOZS-IP-2020-2) from the University Department for Health Studies, University of Split, Split, Croatia.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval The research was approved by the Ethics Committee of the University of Split School of Medicine (class: 003-05/20-03/0001, reg. no.: 2181-198-01-08-20-0040).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Extra data can be accessed via the Dryad data repository at <http://datadryad.org/> with the doi: 10.5061/dryad.v15dv41wd.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Ana Marusic <http://orcid.org/0000-0001-6272-0917>

Tina Poklepović Peričić <http://orcid.org/0000-0002-9686-5062>

REFERENCES

- 1 Rideout VJ, Foehr UG, Roberts DF. *Generation M 2: media in the lives of 8-to 18-year-olds*. Kaiser Family Foundation., 2010.
- 2 Nsangi A, Semakula D, Oxman AD, *et al*. Teaching children in low-income countries to assess claims about treatment effects: prioritization of key concepts. *J Evid Based Med* 2015;8:173–80.
- 3 Manganello JA. Health literacy and adolescents: a framework and agenda for future research. *Health Educ Res* 2008;23:840–7.
- 4 Ettel G, Nathanson I, Ettel D, *et al*. How do adolescents access health information? And do they ask their physicians? *Perm J* 2012;16:35.
- 5 Fergie G, Hunt K, Hilton S. What young people want from health-related online resources: a focus group study. *J Youth Stud* 2013;16:579–96.
- 6 Skinner H, Biscope S, Poland B, *et al*. How adolescents use technology for health information: implications for health professionals from focus group studies. *J Med Internet Res* 2003;5:e32.
- 7 U.S Department of Health and Human Services. *Healthy people 2010: understanding and improving health*. 2nd edn. Washington DC: US Government Printing Office, 2010.
- 8 Betz CL. Health literacy: the missing link in the provision of health care for children and their families. *J Pediatr Nurs* 2007;22:257–60.
- 9 Wolf MS, Feinglass J, Thompson J, *et al*. In search of 'low health literacy': threshold vs. gradient effect of literacy on health status and mortality. *Soc Sci Med* 2010;70:1335–41.
- 10 Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. *Health Promot Int* 2000;15:259–67.
- 11 Coulter A, Ellins J. *Patient-focused interventions: a review of the evidence*. London: Health Foundation, 2006.
- 12 Oxman AD, Oxman M, Semakula D, *et al*. Prioritising and ordering informed health choices (IHC) key concepts to be included in a secondary school resources: protocol. IHC working paper 2019.
- 13 Aronson JK, Barends E, Boruch R, *et al*. Key concepts for making informed choices. *Nature* 2019;572:303–6.
- 14 Oxman AD, Chalmers I, Austvoll-Dahlgren A and informed health choices group. key concepts for assessing claims about treatment effects and making well-informed treatment choices. *F1000Research* 1784;2019:7.
- 15 Austvoll-Dahlgren A, Semakula D, Nsangi A. Measuring ability to assess claims about treatment effects: the development of the "Claim Evaluation Tools". *BMJ Open* 2016;6:e013184.

- 16 Agencija za odgoj i obrazovanje, Ministarstva znanosti, obrazovanja i sporta. Zdravstveni odgoj - priručnik za nastavnike i suradnike u srednjoj školi. Zagreb: MZOS, 2013. Available: <https://mzo.gov.hr/UserDocsImages/dokumenti/Obrazovanje//Priru%C4%8Dnik%20za%20nastavnike%20i%20stru%C4%8Dne%20suradnike%20u%20srednjoj%20%C5%A1koli%20-%20Zdravstveni%20odgoj.pdf> [Accessed 20 Dec 2020].
- 17 von Elm E, Altman DG, Egger M, *et al.* The strengthening of reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol* 2008;61:344–9.
- 18 Austvoll-Dahlgren A, Guttersrud Øystein, Nsangi A, *et al.* Measuring ability to assess claims about treatment effects: a latent trait analysis of items from the claim evaluation tools database using Rasch modelling. *BMJ Open* 2017;7:e013185.
- 19 Bichi AA. Classical test theory: an introduction to linear modeling approach to test and item analysis. *Int J Soc Stud* 2016;2:27–33.
- 20 Matlock-Hetzel S. Basic concepts in item and test analysis. paper presented at the annual meeting of the southwest educational research association. Austin. 23-25 January 1997.
- 21 Adegoke BA. Comparison of item statistics of physics achievement test using classical test and item response theory frameworks. *J Educat Pract* 2013;4:87–96.
- 22 Austvoll-Dahlgren A, Oxman AD, Chalmers I. *The claim evaluation tools Working Group. manual for preparing a test or questionnaire based on the claim evaluation tools item bank*, 2020.
- 23 Bissonnette-Maheux V, Provencher V, Lapointe A, *et al.* Exploring women's beliefs and perceptions about healthy eating blogs: a qualitative study. *J Med Internet Res* 2015;17:e87.
- 24 Abel EK. 'Man, woman, and chore boy': transformations in the antagonistic demands of work and care on women in the nineteenth and twentieth centuries. *Milbank Q* 1995;73:187–211.
- 25 Berg JA, Woods NF. Global women's health: a spotlight on caregiving. *Nurs Clin North Am* 2009;44:375–84.
- 26 Rababah JA, Al-Hammouri MM, Drew BL, *et al.* Health literacy: exploring disparities among college students. *BMC Public Health* 2019;19:1401.
- 27 Oxman AD, Austvoll-Dahlgren A, Garratt A. *Understanding of key concepts relevant to assessing claims about treatment effects: a survey of Norwegian adults. IHC working paper*, 2017.
- 28 Dahlgren A, Furuseth-Olsen K, Rose CJ. The Norwegian public's ability to assess treatment claims: results of a cross-sectional study of critical health literacy [version 1; peer review: awaiting peer review]. *F1000Research* 2020;9:179.
- 29 Pérez-Gaxiola G, Austvoll-Dahlgren A. Validación de un cuestionario para medir La habilidad de la población general para evaluar afirmaciones acerca de tratamientos médicos. *Gac Med Mex* 2018;154:480–95.
- 30 Wang Q, Austvoll-Dahlgren A, Zhang J, *et al.* Evaluating people's ability to assess treatment claims: validating a test in mandarin from claim evaluation tools database. *J Evid Based Med* 2019;12:140–6.
- 31 Austvoll-Dahlgren A, Oxman AD, Chalmers I, *et al.* Key concepts that people need to understand to assess claims about treatment effects. *J Evid Based Med* 2015;8:112–25.
- 32 George D. *SPSS for windows step by step: a simple study guide and reference, 17.0 update, 10/e*. Pearson Education India, 2011.
- 33 Ntoumanis N. *A step-by-step guide to SPSS for sport and exercise studies: a step-by-step guide for students*. Routledge, 2003.