Development and validation of the Conceptual Survey on Wave Optics

Karolina Matejak Cvenic,[1] Maja Planinic,[1,*] Ana Susac,[2] Lana Ivanjek,[3] Katarina Jelicic,[1] and Martin Hopf[4]

[1]*Department of Physics, Faculty of Science, University of Zagreb, Bijenicka 32, HR—10000, Zagreb, Croatia*
[2]*Department of Applied Physics, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, 10000 Zagreb, Croatia*
[3]*Faculty of Physics, Physics Education Research, Technische Universität Dresden, Haeckelstraße 3,01069 Dresden, Germany*
[4]*University of Vienna, Austrian Educational Competence Centre Physics, Porzellangasse 4/2/2, 1090 Wien, Austria*

(Received 11 June 2021; accepted 18 November 2021; published 10 January 2022)

A new diagnostic instrument, the Conceptual Survey on Wave Optics (CSWO), was developed and validated on 224 high school students (aged 18–19 years) in Croatia. The process of test construction, which included the testing of 61 items on the total of 712 students is presented. The final version of the test consists of 26 multiple-choice items which cover basic conceptual aspects of interference, diffraction, and polarization of light at the high school level. The construction of the test and the evaluation of its functioning and validity were conducted using the Rasch model. The test construct was based on the increasing levels of cognitive complexity in accordance with the Webb's depth of knowledge model. The theoretical construct was empirically confirmed, and it suggests that the explanations of wave optics phenomena and their application to real-life situations are the most difficult aspects of high school students' understanding of wave optics. The results of the Rasch analysis suggest that the CSWO is a reliable diagnostic instrument suitable for administration as a post-test to high school students after instruction on wave optics and possibly also for university level students in some introductory physics courses.

DOI: 10.1103/PhysRevPhysEducRes.18.010103

I. INTRODUCTION

Wave optics is an important but difficult topic for high school and university students [1–21]. Previous research on wave optics was conducted mainly at the university level [1–6,10,11,13–17,20,21], but also at the high school level [7–9,12,18,19]. The findings suggest that students have many difficulties with the basic aspects of fundamental wave optics topics, such as interference, diffraction, and polarization of light. In Croatia, where physics is a compulsory subject in many high schools, these topics are taught in the final high school year, when students are 18–19 years old. They are covered typically in about ten 45-min teaching periods, usually through standard lectures with some demonstrations performed by the teacher.

*Corresponding author
maja.planinic@phy.hr

Since most of the authors of this paper are involved in a research project that aims at developing an inquiry-based high school teaching sequence on wave optics, including several students' hands-on investigative experiments, and at assessing its effect on students' conceptual understanding and scientific reasoning, there was a need for a diagnostic instrument that would be suitable for evaluating high school students' conceptual understanding of wave optics. Some instruments on wave optics already exist [6,12,13,16,21], but are not suitable for our purposes because they do not cover the relevant topics or are not applicable at high school physics level. That provided motivation to develop a new diagnostic instrument on wave optics that would meet those demands. We also hope that the new instrument may help other researchers and teachers in the field of wave optics, either as a post-test in high school physics, or maybe also as a pretest, or even a post-test, for wave optics in some introductory physics courses at university level.

The most common problems of students at high school and/or university level, identified in physics education research studies in different countries, can be summarized as follows:

(i) difficulties applying the wave model of light, representing light with wavefronts, and expressing distances in wavelengths of light [1–3,7,10,20]
(ii) difficulties understanding the role of path length difference and correctly applying the interference condition [2,3,9,20]
(iii) using geometrical optics instead of wave optics for explaining wave optics phenomena, especially diffraction on a narrow slit [2,3,7,9,11,17]
(iv) expecting diffraction pattern on a wide slit [3,9]
(v) using mixed models to explain diffraction, e.g., geometrical optics for the middle part of the slit, and wave optics for its edges [3,7,9]
(vi) confusing a polarizer with a narrow slit or an optical grating [3,8,22]
(vii) misinterpreting schematic representations of polarization and confusing the direction of electric field oscillations with the direction of light propagation [8]
(viii) explaining polarization and the functioning of polarizers using geometrical optics [8]
(ix) not being able to distinguish or predict basic wave optics patterns [9,11,18,19]
(x) difficulty explaining the basic wave optics phenomena [3,4,8,9,11,14,15,17]

All these difficulties significantly hinder students' understanding of wave optics, and it would be important for physics teachers to be able to assess it with a diagnostic instrument on wave optics.

In the construction of the CSWO, we were led by the previous research in probing student understanding of the selected topics. The multiple-choice format of the test was chosen because of its advantages for administration and grading. This format is often criticized, however, for not probing student reasoning. We decided therefore to include elements of reasoning in the multiple-choice options on most items, to be able to probe student reasoning on those items.

The test was designed to investigate the following learning outcomes:

1. Demonstrate knowledge of basic wave concepts and the wave model of light
2. Apply mathematical conditions for interference of light from two sources
3. Reason about school experiments in wave optics (interference of light on two slits, interference on an optical grating, diffraction of light on a single slit, and polarization of light with polarizers or by reflection on different media)
4. Differentiate patterns of basic interference and diffraction phenomena introduced in high school physics
5. Explain wave optics phenomena and apply them to real-life situations.

The instructional content of wave optics at high school level that is necessary for students to be able to solve CSWO should include the following topics:

(1) introduction and discussion of the problem of nature of light (particle and wave model of light)
(2) introduction of Young's double slit experiment (for the case when light is incident perpendicularly on the slits), discussion of the obtained pattern in terms of analogy with mechanical wave fronts and of the changes in pattern resulting from changes of light wavelength, slit separation, and slits-screen distance
(3) discussion of coherent and incoherent sources and of the experimental ways of producing coherent sources
(4) introduction, discussion, and application of the mathematical criterion of constructive and destructive interference, application of the concepts of path length and path difference expressed in terms of wavelength
(5) introduction of the optical grating and analysis of the obtained monochromatic interference pattern—comparison with the double slit monochromatic pattern
(6) introduction, discussion, and application of the mathematical expression for the constructive interference condition on optical grating
(7) demonstration and analysis of white light interference pattern on an optical grating
(8) qualitative discussion of examples of interference patterns in nature and everyday life
(9) introduction and analysis of a monochromatic diffraction pattern on a single slit (without introducing the mathematical expression)
(10) demonstration and analysis of the experiment with two polarizers, introduction of the polarization of light and the concepts of polarized and unpolarized light, modeling light as a transverse EM wave
(11) introduction of polarization by reflection and Brewster's law
(12) qualitative discussion of the use of polarization of light in everyday life (e.g., polarizing sunglasses, screens, etc.)

II. APPLICATION OF THE RASCH MODEL TO THE CSWO CONSTRUCTION AND EVALUATION

The construction and evaluation of the CSWO were guided by the Rasch model [23,24]. The process and the most important aspects of test construction and evaluation with the Rasch model are described, for example, by Liu [25] and Planinic *et al.* [26]. The general steps suggested by Liu [25] are the following:

Step 1: Define the construct that can be characterized by a linear attribute.
Step 2: Identify the behaviors corresponding to different levels of the defined construct.
Step 3: Define the outcome space of behaviors (item pool).
Step 4: Field test with a representative sample of the target population.

Step 5: Conduct Rasch modeling.
Step 6: Review item fit statistics and revise items if necessary.
Step 7: Review the Wright map and add or delete items if necessary.
Step 8: Repeat steps 4–7 until a set of items fit the Rasch model and define a scale.
Step 9: Establish validity and reliability claims for the measurement instrument.
Step 10: Develop documentation for the measurement instrument.

The CSWO construction started with the definition of the construct, which was student understanding of basic phenomena of wave optics at high school level. The construct was designed to show progression according to the chosen cognitive model, in our case Webb's depth of knowledge (WDK) model [27]. In that model there are three levels of knowledge that are accessible to examination with a written test, which can be labeled as

Level 1. Recall and reproduction
Level 2. Skills and concepts
Level 3. Strategic thinking

Webb's level 4, "extended thinking" can be developed and tested through student projects and longer investigations but is not suitable for probing with tests. The Rasch model requires that the construct (in this case student understanding of wave optics) is organized hierarchically, meaning that its items should form a certain "ladder" of difficulties, which should ideally not change from one group of students to the other (provided they are at the same level of learning and have covered in teaching the topics that are probed in the test). Such hierarchy is not easy to achieve in conceptual tests since each item difficulty may depend on and vary with the particular focus and emphasis of instruction. We therefore decided to organize the construct around cognitive complexity of items in accordance with the WDK model. The WDK model is easier to apply than most cognitive models because it relates cognitive level to item complexity (higher-order cognitive processes require more steps and cognitive resources than lower-order processes) and increasing complexity is relatively easy to recognize in items.

We have therefore built in our construct the progression from level 1 (knowledge about basic wave concepts, such as wavelength of light or path difference, as well as knowledge about wave model of light, that are both needed for building of understanding of more complex phenomena, such as interference, diffraction or polarization), through level 2 (being able to recognize important features and variables of each phenomenon and reason about their changes in typical school experiments) to level 3 (being able to give explanations of different phenomena and apply them to real-life situations). With this, we have defined the construct and the behaviors that correspond to its different levels (steps 1 and 2). Based on previous research, the literature on wave optics, and the semistructured demonstration interviews conducted with high school students [8,9] we have developed an initial set of 38 items (step 3). Some of these items were tested in open-ended format on a sample of ca. 100 high school students, on two occasions during the test development. Student answers, together with the results from the interviews, helped to create or improve some distracters on the multiple-choice items. After having formed the final version of the multiple-choice items, the field testing of the first item set was conducted (step 4) and Rasch analyzed with Winsteps (step 5). Since 38 items were too many for students to solve in 45 min, we split the set into two 24-item versions, each containing 10 common items. About half of the student sample was given the first, and the other half the second version, so each student was solving only 24 items, but the common items enabled the linking of the two parts and the analysis of the whole item set together. The fit of data to the model was evaluated and the fit statistics and the point-measure correlations inspected (step 6). Poorly fitting items were either removed from the instrument or sometimes reformulated and tested in the later versions of the CSWO. The Wright item-person map was inspected, and some new items were introduced (step 7). The steps 4–7 were repeated in three cycles (involving the total of 61 tested items and 712 students, see Table I[1]) until a set of 26 well-functioning items was obtained (step 8). The final version of the instrument was validated through inspection of its Rasch parameters (Infit MNSQ, Outfit MNSQ, point-measure correlations, item, and person reliability indices) and the item-person map, bubble chart, ICC curves, DIF, and PCA analysis. The empirical construct was then compared with the theoretical construct (step 9). The process of test development described in this paper can be considered a part of its documentation (step 10).

We will shortly describe here theoretically the Rasch parameters and techniques mentioned above. Their values for the CSWO will be presented and discussed in Sec. IV. A complete introduction to the Rasch model for readers not familiar with it can be found in Refs. [23,28] and a short introduction in Refs. [26,29]. The Rasch analysis is based on the analysis of the fit of data with the Rasch model, which is performed through the comparison of the differences between the theoretical and the experimental values (residuals) for both students and items. Commonly used fit statistics are Outfit MNSQ (mean squares) statistics, as the arithmetic mean of simple squared residuals, and Infit MNSQ as a weighted mean of squared residuals. Outfit is more sensitive to outliers, and infit to respondent's responses to items whose difficulties are close to respondent's ability. Good model fit is characterized by Infit and Outfit MNSQ values for items close to the model value

[1]All CSWO versions, except the 4th version, were tested on Croatian students. The CSWO 4 was tested on high school students in Vienna, Austria.

TABLE I. Overview of the process of the CSWO development.

CSWO version	Time of testing	Number of items	Number of participants	Type of students	Items discarded after Rasch analysis	New items added after Rasch analysis
CSWO 1 and 2	April 2019	24 items each (10 common items, so 38 items in total)	139 and 143	High school students (aged 18–19)	15	15
CSWO 3	May/June 2019	38	47	University students (aged 22–24)	5	0
CSWO 4	September 2019	33	61	High school students (aged 17–19)	7	4
CSWO 5	October 2019	30	98	University students (aged 19–20)	4	4
CSWO 6	October 2019–February 2020	30	224	High school students (aged 18–19)	4	0

of 1, in practice usually between 0.7 and 1.3 [25]. Although items with Infit and Outfit MNSQ values in a broader range, between 0.5 and 1.5, can be acceptable and not degrading for measurement [30], such items still need to be inspected for the reasons of their larger misfit. Point-measure correlations help us to see how a specific item contributes to the whole person or item measure. A bubble chart [23] can help visualize the overall structure and functioning of the test and its items. It is a graphical presentation of item difficulty vs item Outfit MNSQ or Infit MNSQ, representing each item as a circle, whose size is proportional to its standard error (smaller circles represent more precisely calibrated items).

Poor fit of some items may be a sign of problems with item wording, scoring, or content. Persons (students) can also show misfit, which may signal problems with student behavior, such as guessing or carelessness during test solving. In the process of test development, such students need to be identified and removed from the analysis, because they create noise in the data and may cause unnecessary removal of good items which may therefore appear as misfitting.

The structure of the test is examined with the use of the Wright map (item-person map), which presents both item difficulties and student abilities (not in the sense of any general student ability, such as, e.g., intelligence, but as a measure of the amount of the investigated latent trait) along the same logit scale. Using the Wright map it is easy to visualize the targeting of the test to the sample, as well as the targeting of individual items to persons, and compare the width of the distribution of test items and the width of the target population ability distribution. Large gaps between the item difficulties in the test or the unnecessary crowding of items in some regions signal problems in the test structure.

To check the dimensionality of the test (unidimensionality is one of the basic requirements of the Rasch model), point-measure correlations and item fit should be inspected and the principal component analysis (PCA) of residuals may be performed. Item misfit and/or strong correlations among residuals may be signs of other dimensions in the test.

When a satisfactory set of items is established, and the instrument shows good functioning, the invariance properties of both the person and item measures can be inspected, for example, through scatter plots of item measures of different subgroups of examinees. For a well-constructed instrument, these measures should be essentially the same, within the limits of their standard errors. If an item in the scatter plot departs significantly from the identity line, it exhibits differential item functioning (DIF), and it should be further examined, revised, or possibly removed from the test.

The validity of the constructed instrument is evaluated with the Rasch model both theoretically and empirically. The theoretical validity is already built in through the construction process requiring a definition of the variable (steps 1 and 2 described above). However, it is the chosen items that operationally define the construct, so their face validity must be checked by the experts in the field. The empirical validity check is conducted by estimating how well the chosen items have succeeded in defining a sufficiently unidimensional and internally coherent construct. This can be investigated through the analysis of item fit, item correlations, and test unidimensionality.

Test reliability can be understood as the degree of the reproducibility of measures. This can be checked with Rasch person reliability (analogous to the classical Cronbach alpha index), but also with item reliability, an index without a similar classical analog. Both reliabilities can take values between 0 and 1. Person reliability of 0.5 is considered to be the minimum meaningful reliability [30]. High reliability may signal good reproducibility of the results but does not necessarily imply that the quality of the test is also high. High quality of the test starts with good items and is reflected in good structure and functioning of the test and its individual items.

III. METHODOLOGY

A. The interviews

The process of the CSWO construction started with a qualitative study that had a goal to determine students' difficulties with basic wave optics phenomena, such as interference, diffraction, and polarization of light. It included 27 semistructured demonstration interviews with Croatian high school students (aged 18–19 years) after their regular school instruction on wave optics. That instruction included all topics that were probed during the interviews, as well as most of the demonstrations that were presented to students in the interviews. The interviews probed student reasoning about school experiments related to these topics. During the interviews, students were shown four standard experiments (interference of light on a double slit and on an optical grating, diffraction of light on a single slit, and polarization of light with different orientation of two polarizers) and were asked for their predictions, observations, and explanations. The findings provided information about students' understanding of these topics and the related experiments, which was in most cases rather poor. The detailed overview and analysis of the findings from the interviews regarding polarization of light can be found in Ref. [8], preliminary findings on diffraction and interference in Ref. [9], while the detailed overview and analysis of the latter findings are currently in preparation. The same interviews were simultaneously conducted in Austria during the spring of 2018, with six high school (gymnasium) students in Vienna, after their regular instruction on wave optics, and the findings were overall similar to those obtained in interviews with Croatian students [9].

When reasoning about polarization of light, the interviewed students very often based their answers on misinterpretation of the schematic representations and the analogies typically used in teaching on polarization, such as mistaking the different oscillation directions for the light propagation directions or concluding that slits are polarizers [8]. Students showed problems with predicting and differentiating patterns of interference and diffraction of light, which may be due to inadequate observations of these patterns and the lack of students' hands-on experiments in typical physics teaching on wave optics. The similar problems with pattern differentiation and prediction were found using the eye-tracking technique on other high school respondents in two additionally conducted studies [18,19]. Findings regarding student difficulties with interference and diffraction were mostly in line with the already known difficulties from other studies [1–20], as summarized in the Introduction. Generally, students had problems already at the level of recognizing and describing phenomena, but even more with explaining them or applying them to some real-life situations. The findings suggested that students had not formed models of these phenomena during instruction, and that they often seemed to revert to the fragments of factual knowledge, coming from either geometrical or wave optics, and combined them on the spot to produce some explanations, but mostly inadequate [8].

B. CSWO development and sampling

Based on the interviews and findings from other physics education research studies, the literature on wave optics, and high school physics curricula the learning outcomes were formulated and an initial pool of items (in Croatian language) for the CSWO was generated. The observed student difficulties from the interviews were used for developing some of the distracters for the CSWO items. An example of a CSWO item is provided in Fig. 1 and the English translation of the whole test is given in the Supplemental Material [31].

The distracters in this item are based on students' explanations provided for the optical grating pattern in the interviews. Students often used geometrical optics to explain the observed pattern. For example, some used the reflection or refraction of light when explaining optical grating pattern, and some expected to see as many maxima on the screen as there were slits.

The process of the CSWO construction included developing and testing several versions of the CSWO, as presented in Table I. The CSWO was developed primarily for assessing high school students' understanding of basic wave optics concepts after school instructions, but we believe that it could also be used with some university students. This is one of the reasons why both high school and university students participated in the CSWO testing (Table I). The type of students differed also because the authors' access to high school students who had completed wave optics was not always possible, so samples of university students were sometimes used instead. The aim of the early testing was to eliminate poorly functioning items, which was possible to achieve also with university students—if an item showed poor functioning (large misfit) in the university population, it would certainly be poorly functioning for high school students too. The reverse was not automatically true, but since all the remaining items would be tested on high school students in the last step,

Q22 Why do we observe alternating minima and maxima of light on the screen after passing laser light through an optical grating?

A. On an optical grating light is reflected at different angles, and that is why we see minima and maxima at different positions on the screen.
B. Each slit of an optical grating acts as a point source, so we see on the screen an interference pattern created by all those sources, which consists of minima and maxima.
C. Each slit on the optical grating lets light through and produces its own maximum, while light is blocked between the slits, forming minima.
D. On an optical grating light is refracted at different angles, and that is why maxima are formed only at certain angles, while minima are between them.

FIG. 1. Item Q22 from the CSWO.

such problems could still have been spotted and remedied at that point. To check whether the test could function in a different educational setting from Croatian, we tested one version of the test on Austrian high school students, after they had finished instruction on wave optics.

Development of the CSWO included several cycles of testing and Rasch analysis. As presented in Table I, after each cycle of testing and Rasch analysis, some items were discarded, and some new items were added. Altogether we tested 61 items, involving 712 students in total. Some of the new tested items were revisions of old items (with, e.g., improved wording). Items were inspected by several experienced physics teachers to confirm their face validity.

The CSWO 6 was administered in Croatia, during the school year of 2019/20, before the outbreak of the COVID-19 pandemic. For this cycle of testing, we needed around 200 high school students, who could take the test right after their regular school instructions on wave optics, but before their regular school test on this topic. Through high school physics teachers, who were interested in participating in our research project, we obtained the sample of 224 gymnasium students (gymnasium is a type of high school in Croatia that typically prepares students for continuation of education at universities). Before taking the test, students had spent 4–5 weeks covering wave optics, with two or three 45-min lessons per week (depending on the type of gymnasium), mostly lecture based. In this sample, there were 145 female and 77 male students, which reflects the typical gender distribution in Croatian gymnasiums. Two students did not specify their gender. The test was administered on paper, in Croatian language, and the allocated time for solving the test was 45 min, but most students finished it in about 30 min. Students solved it anonymously, using a code name, and were not given any incentive for solving the test, such as grades, but were informed of their result later (under code names).

After Rasch analysis of the CSWO 6, four items were discarded due to misfit, leaving the final version of the CSWO with 26 items, distributed across learning outcomes in the following way:

1. Demonstrate knowledge of basic wave concepts and the wave model of light (Q1, Q5, Q6, Q8, Q25, Q26)
2. Apply mathematical condition for interference of light from two sources (Q3, Q4, Q13)
3. Reason about school experiments in wave optics (Q7, Q10, Q11, Q12, Q14, Q17, Q19, Q21, Q24)
4. Differentiate patterns of basic interference and diffraction phenomena introduced in high school physics (Q15, Q18, Q23)
5. Explain wave optics phenomena and apply them to real-life situations (Q2, Q9, Q16, Q20, Q22).

The third learning outcome was probed with the largest number of items, because there were three wave optics phenomena that needed to be included and each was represented by several items. Covering different phenomena was also the reason for a larger number of items in the first and the fifth outcome, whereas the second and the fourth outcome could be probed by only three items each, because there are only three different patterns that high school students should differentiate (outcome 4) and the applying of the mathematical condition for interference of light is a single procedure or skill that does not require more items at the basic high school level (outcome 2). However, we believe that this procedure is important, as it underlies explanations of many wave optics phenomena, so we singled it out as a separate outcome. The pattern differentiation was singled out as a separate outcome since we noticed in the interviews and the earlier eye-tracking studies that this was unexpectedly difficult for students, so we wanted to be able to diagnose this difficulty with the CSWO, if it appeared.

To inspect empirically the proposed theoretical construct of the CSWO, we have performed an additional Rasch analysis of the data with five groups of items (determined by the five learning outcomes). Student raw scores for each group of items were added up and each group was then analyzed as a new separate "item." Each item group had a different maximum score, determined by the number of items in the group, which required Rasch analysis with a partial credit model [23,30]. The obtained difficulties of item groups were compared, and it was checked whether these item groups represented different difficulty strata in the theoretical construct of the test.

IV. RESULTS AND DISCUSSION

The results of the Rasch analysis of the 26 items in the last version of the CSWO, administered to 224 Croatian high school (gymnasium) students, will be presented in this section. General principles of Rasch analysis and the meanings of its output are explained in more detail, for readers not familiar with it, in other publications [23,26,28,29], so here we will only discuss the results of the present analysis. The test and the frequencies of student answers for each question and each distractor are given in the Supplemental Material [31]. The CSWO was analyzed with Winsteps Rasch software [30,32]. In Fig. 2 the Wright map (item-person map), and in Fig. 3 the bubble chart for the CSWO is shown. Both figures visualize the structure of the test, with the emphasis on the comparison of the distribution of items with the distribution of students in item-person map and the emphasis on item fit in the bubble chart. The vertical line in the middle of the item-person map represents the underlying variable in the test (student understanding of wave optics) with the logit scale along which are placed both students and items to allow visual comparison of the two distributions and estimation of targeting and width of the test. The item-person map shows good targeting of the test on the sample. The test has good width, covering almost all respondents and good distribution of items according to their difficulty. Some lack of easier items is noticeable, but is not critical, since all students have items in the ± 1 logit range around their ability

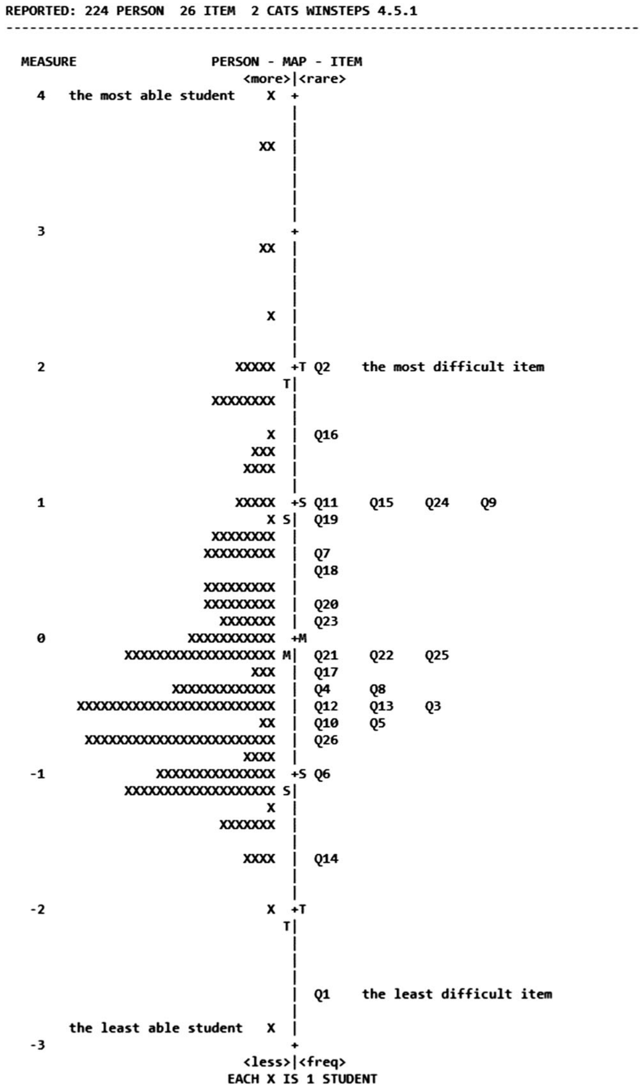

FIG. 2. Wright map (item-person map) for the CSWO. Each "x" represents one student. The distribution of students on the left-hand side and the distribution of the CSWO items on the right-hand side are shown along the same logit axis in the middle, where M stands for the mean of each distribution, and S and T mark points 1 or 2 standard deviations away from the mean.

estimate, as is required for good measurement [30,33]. On the other hand, for the top three students, this requirement is not fulfilled, and the test could be supplemented by at least one difficult item around the level of 3 logits or more to match those students. However, since these few students are very far away in ability from the rest of the students, and since for most of the students the test is already quite difficult, it was concluded that adding more difficult items was not necessary at this point. In the middle of the test some items appear close together, however, they are not redundant since they test different phenomena of wave optics. The test length is adequate for the allocated testing time and the Rasch analysis requirements (reducing the number of items would result in larger uncertainties of person measures).

The fit of items with the model is very good, which is shown in the bubble chart in Fig. 3, where the allowed range of fit (0.5–1.5 logit) is displayed. The bubble chart enables visual inspection of the fit of items—items that are close to the expected value of fit of 1 are close to the central line in the figure. It also shows the uncertainties in item difficulties, indicated by the size of the bubbles.

All items are not only well within that range, but also quite close to the central line which represents the model value of 1 for fit. In Table II.[2] the values of item misfit (Infit and Outfit MNSQ values) and point measure correlations can be found for each item. All misfit values are within the recommended 0.7–1.3 range (the only exception is the Outfit value of item Q20 which is 1.32), so it can be concluded that the test items show good fit with the Rasch model expectations. The inspection of point-measure correlations shows no items with negative or very low correlations.

The empirical unidimensionality of the construct was inspected through the principal component analysis that is a part of the regular Winsteps output [30]. The values of all contrasts had strengths below 2, suggesting that if another dimension existed in the test, it could be identified by less than two items, which is not enough to consider it a separate dimension [30]. To investigate possible differential item functioning (DIF) for male and female students, the scatter plot of item difficulty measures obtained for these two subsamples was created and is presented in Fig. 4. A point on that plot outside the confidence bands, which are defined by the standard errors of the item calibrations, would suggest an item that showed significantly different functioning for students of different gender, and such item would have to be removed from the test. Figure 4 suggests that all the CSWO items are free of significant DIF, therefore confirming the CSWO construct invariance for these subsamples of students. There are three borderline items, Q9, Q20, and Q24, but their DIF is still not critical.

The test has an acceptable Cronbach alpha measure of 0.77, which is reflected also in the Rasch person reliability index of 0.78. Since the test is not designed to be used for any decision making about students (which would require person reliabilities of 0.8 or more [30]), but as a diagnostic tool only, this is an acceptable value. The item reliability is quite high (0.97) and close to the maximum value of 1. The difference in the values of person and item reliabilities is mostly due to the difference in the number of items and persons.

Overall, the analysis suggests that the chosen item set has succeeded in creating a sufficiently coherent and unidimensional construct from the Rasch model perspective, with good validity and reliability.

[2]During the review process of the article, it was pointed out to us by one of the reviewers that the wordings of questions Q3 and Q4 needed additional specification, and we added that the sources were "coherent and in phase" (Q3) and "in phase" (Q4) (Supplemental Material [31]). The difficulties of Q3 and Q4 refer to the wordings of these questions without those additions.

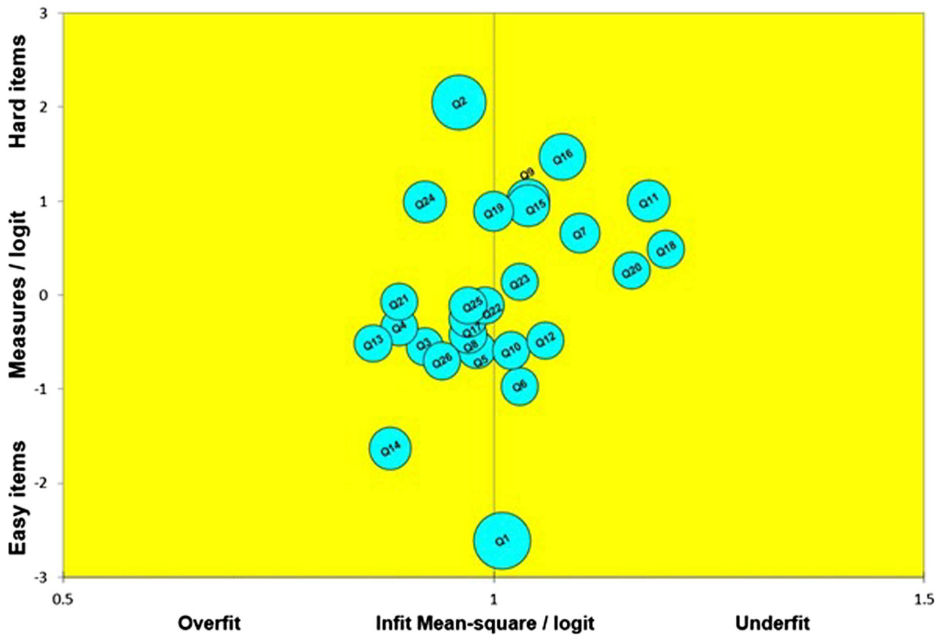

FIG. 3. Bubble chart of the CSWO item difficulties vs Infit MNSQ values in logit. Each item is represented by a circle whose size is proportional to its calibration uncertainty (standard error). The maximum range of item misfit that is acceptable for measurement is shown (0.5–1.5 logit). The central line represents the model value of Infit of 1.

To inspect empirically the proposed theoretical construct of the CSWO, we have conducted the partial credit analysis, with its results presented in Table III.

TABLE II. Percentages of students' correct answers, item difficulties, standard errors, Infit MNSQ and Outfit MNSQ (all four in logit), and point-measure correlations.

Item	Correct answers (%)	Item difficulty	Standard error	Infit MNSQ	Outfit MNSQ	Point-measure corr
Q1	89.7	−2.61	0.23	1.01	1.00	0.18
Q2	13.4	2.05	0.22	0.96	1.01	0.43
Q3	56.7	−0.55	0.15	0.92	0.88	0.44
Q4	52.2	−0.34	0.15	0.89	0.83	0.49
Q5	57.6	−0.58	0.15	0.98	0.96	0.39
Q6	66.1	−0.97	0.15	1.03	0.96	0.32
Q7	32.6	0.66	0.16	1.10	1.25	0.33
Q8	54.4	−0.42	0.15	0.97	0.92	0.41
Q9	27.2	1.01	0.17	1.04	1.07	0.39
Q10	58.0	−0.59	0.15	1.02	0.99	0.35
Q11	27.4	1.00	0.17	1.18	1.18	0.28
Q12	56.3	−0.48	0.15	1.06	1.05	0.32
Q13	56.3	−0.51	0.15	0.86	0.80	0.50
Q14	77.2	−1.63	0.17	0.88	0.71	0.41
Q15	28.1	0.95	0.17	1.04	1.02	0.40
Q16	19.6	1.47	0.19	1.08	1.12	0.33
Q17	50.4	−0.25	0.15	0.97	0.91	0.42
Q18	36.2	0.49	0.15	1.20	1.23	0.25
Q19	29.0	0.89	0.16	1.00	0.99	0.43
Q20	40.2	0.26	0.15	1.16	1.32	0.25
Q21	46.4	−0.07	0.15	0.89	0.85	0.49
Q22	48.2	−0.11	0.15	0.99	0.95	0.41
Q23	42.9	0.14	0.15	1.03	1.02	0.38
Q24	26.8	0.99	0.17	0.92	0.91	0.49
Q25	46.9	−0.11	0.15	0.97	0.94	0.43
Q26	59.4	−0.70	0.15	0.94	0.90	0.41
MEAN	46.1	0.00	0.16	1.00	0.99	

The resulting difficulty levels are shown in Fig. 5. Since Winsteps sets the average difficulty of items to zero, groups with negative values are easier than the test average, and groups with positive values are more difficult than the test average. Each group is shown with an error bar equal to its three standard errors, and the absence of overlap in error bars suggests that the groups represent statistically distinct difficulty strata at the level of 0.05 significance [34]. Figure 5 suggests that students have the least problems with knowledge about phenomena and basic wave concepts, as well as with mathematical application of the interference condition. Reasoning about experiments in wave optics is of average difficulty but differentiating basic interference and diffraction phenomena according to their characteristic patterns seems to be more difficult, which agrees with our initial assumptions. The finding concerning experimental patterns is confirmed also by our earlier study, in which we used eye-tracking technique to investigate students' ability to predict and recognize different patterns in wave optics [18]. Although it may seem at first that this should be an easy task for students, it turns out to be a rather difficult one. Students often do not notice in typical quick classroom demonstrations the characteristic features of those patterns, so overall they may all seem very similar to them. Also, this is not just a simple task of remembering and recognizing the images, since experimentally obtained patterns may vary quite significantly with the used equipment (for example, they may vary in the degree of presence of the diffraction phenomena on a double-slit experiment, which is meant to demonstrate the ideal interference pattern only), therefore students must know theoretically the main

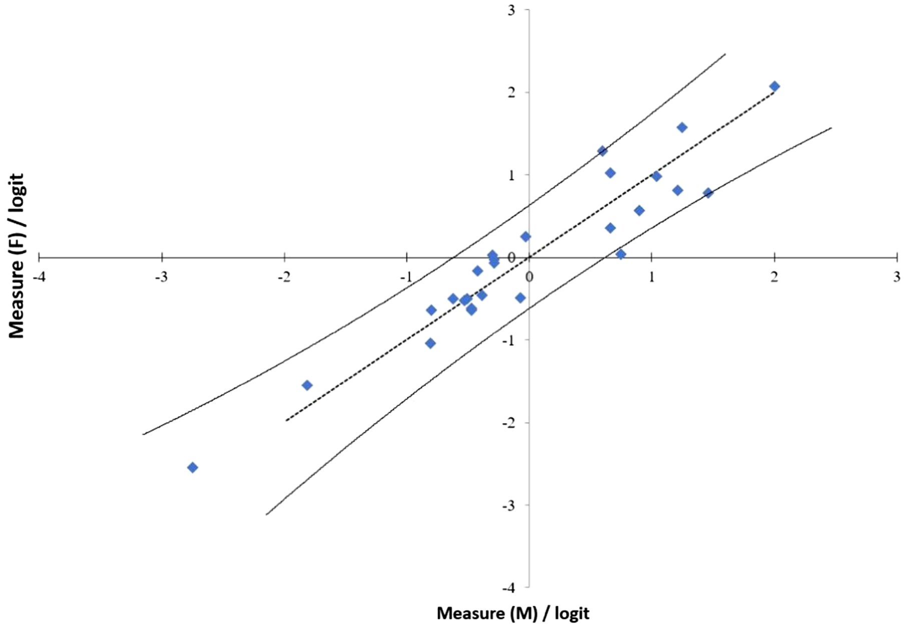

FIG. 4. Scatter plot of item difficulties, obtained from separate analyses of male and female students, with confidence bands.

characteristic features of each pattern, analyze the image, and look for those features to make the correct identification. So, in the end, this task turns out to be cognitively demanding.

The most difficult group of items is the one concerning the explanations and applications of wave optics phenomena. This was expected, since the ability to explain and apply requires that students have a good model of each phenomenon, which does not seem to have been formed by most of the tested students. Instead, students seemed to have combined many fragments of knowledge on those items and chose explanations that were consistent with student difficulties identified in the existing research on student understanding of wave optics (e.g., attempting to explain wave phenomena with geometrical optics). For example, on Q22 in Fig. 1, 19% of the students chose option D (refraction), 10% option A (reflection) and 22% chose option C (each slit produces its own maximum) to explain the formation of the optical grating pattern of maxima and minima.

The observed hierarchy of the difficulties of the investigated groups of items seems to be consistent with the hypothesized theoretical expectations based on the Webb's depth of knowledge model.

Unlike procedural tests in which the hierarchy of items in the construct is relatively easy to define through the increase in the complexity of the procedure (e.g., multiplying two-digit numbers is more difficult than multiplying one-digit numbers), constructs and hierarchy of items in conceptual tests are much more difficult to establish in an objective way. We believe that the principle of increasing cognitive complexity may be a good way to organize conceptual constructs and make them applicable in different educational systems.

In the future, we hope to apply the CSWO to samples of students in other countries and to further check its functioning and its diagnostic power. By now, the German version of the CSWO has been created and administered to students. The English version has not yet been applied. One possible limitation for the application of the test in other countries could be the differences in high school physics curricula, especially in topics covered in wave optics, but we hope that being constructed around the most basic wave

TABLE III. Percentages of students' correct answers, difficulties, standard errors, Infit MNSQ and Outfit MNSQ (all four in logit), and point-measure correlations for each group of items, obtained through partial credit analysis.

Group of items	Correct answers (%)	Difficulty	Standard error	Infit MNSQ	Outfit MNSQ	Point-measure corr
Knowledge	62.1	−1.20	0.07	0.83	0.81	0.75
Interference condition	55.0	−0.54	0.08	0.90	0.92	0.64
Experiments	44.7	0.06	0.06	0.83	0.83	0.83
Patterns	35.6	0.51	0.09	1.31	1.29	0.52
Explanations	29.6	1.17	0.08	1.15	1.14	0.65

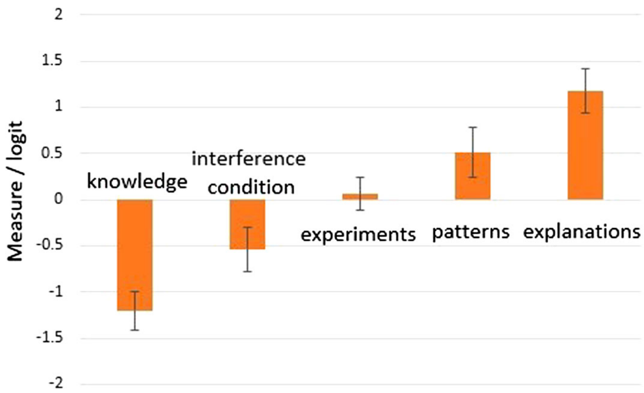

FIG. 5. Empirical validation of the theoretical construct of the CSWO. Difficulties of the five groups of items belonging to the different levels of the construct are shown with error bars indicating three standard errors for each group.

optics phenomena, the CSWO might be applicable in a wide range of educational settings.

V. CONCLUSIONS

The CSWO is a new diagnostic instrument on wave optics, primarily constructed to be a post-test for high school students, but possibly also applicable at the university level as a pretest, to indicate the level of student knowledge before starting instruction on wave optics, or even as a post-test in some university physics courses. The test was constructed and evaluated using the Rasch model. It shows good structure and functioning. The CSWO seems to have identified and operationally defined a construct of high school students' conceptual understanding of wave optics and related it to the different cognitive operations required for solving different wave optics items. The results of Croatian high school students suggest that the most difficult aspect of wave optics understanding is forming explanations and applying knowledge to real-life phenomena. These aspects require that students form good models of phenomena, which does not seem to be often the case in standard teaching. However, it was also noticed that the pattern differentiation, as well as some aspects of reasoning about basic wave optics experiments posed significant problems to students. We suggest that the introduction of students' hands-on experiments in an inquiry-based teaching sequence might help students to better observe, investigate, and differentiate different phenomena and to form better models. The CSWO could be of help to high school teachers and even university faculty as a good diagnostic instrument to indicate the level of student understanding and help improve the teaching and learning of wave optics.

ACKNOWLEDGMENTS

This work has been fully supported by the Croatian Science Foundation under the Project No. IP-2018-01-9085.

[1] A. Coetzee and S. N. Imenda, Alternative conceptions held by first year physics students at a South African university of technology concerning interference and diffraction of waves, Res. High. Educ. J. **16**, 13 (2012), https://www.aabri.com/manuscripts/121097.pdf.

[2] B. S. Ambrose, P. R. L. Heron, S. Vokos, and L. C. McDermott, Student understanding of light as an electromagnetic wave: Relating the formalism to physical phenomena, Am. J. Phys. **67**, 891 (1999).

[3] B. S. Ambrose, P. S. Shaffer, R. N. Steinberg, and L. C. McDermott, An investigation of student understanding of single-slit diffraction and double-slit interference, Am. J. Phys. **67**, 146 (1999).

[4] P. Colin and L. Viennot, Using two models in optics: Students' difficulties and suggestions for teaching, Am. J. Phys. **69**, S36 (2001).

[5] E. M. Colombo, M. Jaen, and L. C. de Cudmani, in *SPIE Proceedings Vol. 2525*, edited by M. J. Soileau (SPIE International Society for Optics and Photonics, Bellingham, WA, 1995), pp. 452–458.

[6] R. Dai, J. C. Fritchman, Q. Liu, Y. Xiao, H. Yu, and L. Bao, Assessment of student understanding on light interference, Phys. Rev. Phys. Educ. Res. **15**, 020134 (2019).

[7] P. Hubber, Year 12 students' mental models of the nature of light, Res. Sci. Educ. **36**, 419 (2006).

[8] K. Matejak Cvenic, L. Ivanjek, M. Planinic, K. Jelicic, A. Susac, and M. Hopf, Analyzing high school students' reasoning about polarization of light, Phys. Rev. Phys. Educ. Res. **17**, 010136 (2021).

[9] K. Matejak Cvenic, L. Ivanjek, M. Planinic, A. Susac, K. Jelicic, M. Hopf, and C. Srnka, Exploring secondary school students' understanding of basic phenomena relating to wave optics, J. Phys. Conf. Ser. **1929**, 012007 (2021).

[10] M. Kryjevskaia, M. R. Stetzer, and P. R. L. Heron, Student difficulties measuring distances in terms of wavelength: Lack of basic skills or failure to transfer?, Phys. Rev. ST Phys. Educ. Res. **9**, 010106 (2013).

[11] L. Maurines, Geometrical reasoning in wave situations: The case of light diffraction and coherent illumination optical imaging, Int. J. Sci. Educ. **32**, 1895 (2010).

[12] V. Mešić, E. Hajder, K. Neumann, and N. Erceg, Comparing different approaches to visualizing light waves: An experimental study on teaching wave optics, Phys. Rev. Phys. Educ. Res. **12**, 010135 (2016).

[13] V. Mešić *et al.*, Measuring students' conceptual understanding of wave optics: A Rasch modeling approach, Phys. Rev. Phys. Educ. Res. **15**, 010115 (2019).

[14] V. Mešić, A. Vidak, E. Hasović, and M. Đekić, University students' ideas about the role of the aperture, and laser beam dimensions in formation of diffraction patterns, Eur. J. Phys. **40**, 055701 (2019).

[15] S. K. Şengören, Turkish students' mental models of light to explain the single slit diffraction, and double slit interference of light: A cross-sectional study, J. Balt. Sci. Educ. **9**, No. 1, 61 (2010), http://oaji.net/articles/2014/987-1404741216.pdf.

[16] D. Salibašić Glamočić, V. Mešić, K. Neumann, A. Sušac, W. J. Boone, I. Aviani, E. Hasović, N. Erceg, R. Repnik, and V. Grubelnik, Maintaining item banks with the Rasch model: An example from wave optics, Phys. Rev. Phys. Educ. Res. **17**, 010105 (2021).

[17] S. K. Şengören, How do Turkish high school graduates use the wave theory of light to explain optics phenomena?, Phys. Educ. **45**, 253 (2010).

[18] A. Susac, M. Planinic, A. Bubic, L. Ivanjek, and M. Palmovic, Student recognition of interference and diffraction patterns: An eye-tracking study, Phys. Rev. Phys. Educ. Res. **16**, 020133 (2020).

[19] A. Susac, M. Planinic, A. Bubic, K. Jelicic, L. Ivanjek, K. Matejak Cvenic, and M. Palmovic, Effect of students' investigative experiments on students' recognition of interference and diffraction patterns: An eye-tracking study, Phys. Rev. Phys. Educ. Res. **17**, 010110 (2021).

[20] K. Wosilait, P. R. L. Heron, P. S. Shaffer, and L. C. McDermott, Addressing student difficulties in applying a wave model to the interference and diffraction of light, Am. J. Phys. **67**, S5 (1999).

[21] M. Roskot, Conceptual Test in Wave Optics, J. Phys. Conf. Ser. **1929**, 012062 (2021).

[22] H. D. Park, K. Kim, M. Kwon, C. Lee, S. Kim, and J. B. Kim, in *Proceedings of the International Conference on Physics Education in Cultural Contexts*, edited by Y. Park and Y. Park (Cheongwon, South Korea, 2001), pp. 190–196.

[23] T. G. Bond and C. M. Fox, *Applying the Rasch Model. Fundamental Measurement in the Human Sciences* (Erlbaum, Mahwah, NJ, 2001).

[24] G. Rasch, *Probabilistic Models for Some Intelligence and Attainment Tests* (Danmarks Paedagogiske Institut, Copenhagen, 1960).

[25] X. Liu, *Using and Developing Measurement Instruments in Science Education. A Rasch Modeling Approach* by Xiufeng Liu (Information Age Pub, Charlotte, NC, 2010).

[26] M. Planinic, W. J. Boone, A. Susac, and L. Ivanjek, Rasch analysis in physics education research: Why measurement matters, Phys. Rev. Phys. Educ. Res. **15**, 020111 (2019).

[27] Norman Webb, Depth-of-knowledge levels for four content areas, 2002, http://ossucurr.pbworks.com/w/file/fetch/49691156/norm%20web%20dok%20by%20subject%20area.pdf.

[28] W. J. Boone, J. R. Staver, and M. S. Yale, *Rasch Analysis in the Human Sciences* (Springer Netherlands, Dordrecht, 2014).

[29] X. Liu and W. J. Boone, *Applications of Rasch Measurement in Science Education* (JAM Press, Maple Grove Minnesota, 2006).

[30] J. M. Linacre, A user's guide to Winsteps Ministep Rasch-model computer programs, 2018, https://www.winsteps.com/a/Winsteps-Manual.pdf.

[31] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.18.010103 for the complete CSWO with the frequencies of students' answers for each question.

[32] J. M. Linacre, Winsteps Rasch tutorials 1–4, https://www.winsteps.com/tutorials.htm, accessed Apr 18, 2021.

[33] Benjamin D. Wright and Mark H. Stone, *Best Test Design* (MESAPress, Chicago, IL, 1979).

[34] J. M. Linacre, Number of Person or Item Strata: $(4 * \text{Separation} + 1)/3$, https://www.rasch.org/rmt/rmt163f.htm.