C.3.4.5 Proficiency classification of Freedom (of expression, of speech, of press, of association/organisation) (socio-emotional)

For the SDG category “Human Rights”, subcategory “Freedom (of expression, of speech, of press, of association/organisation)”, we proposed to consider the responses to the items present in “What is good for democracy” section from ICCS 2016 (W. Schulz, Carstens, et al., 2018). These are presented in the following figure:

Figure 22. Students' reports on students’ opinions regarding what is good for democracy in ICCS 2016

Q22 Below is a list of things that may happen in a democratic country. Some of them may be good for and strengthen democracy, some may be bad for and weaken democracy, while others are neither good nor bad for democracy.

Which of the following situations do you think would be good, neither good nor bad, or bad for democracy?

(Please tick only one box in each row.)

			Good for democracy	Neither good nor bad for democracy	Bad for democracy	
IS3G22A	a)	Political leaders give government jobs to their family members.	☐1	☐2	☐3	td06
IS3G22B	b)	One company or the government owns all newspapers in a country.	☐1	☐2	☐3	td07
IS3G22C	c)	People are allowed to publicly criticize the government.	☐1	☐2	☐3	td01
IS3G22D	d)	All adult citizens have the right to elect their political leaders.	☐1	☐2	☐3	td02
IS3G22E	e)	People are able to protest if they think a law is unfair.	☐1	☐2	☐3	td03
IS3G22F	f)	The police have the right to hold people suspected of threatening national security in jail without trial.	☐1	☐2	☐3	td08
IS3G22G	g)	Differences in income between poor and rich people are small.	☐1	☐2	☐3	td04
IS3G22H	h)	The government influences decisions by courts of justice.	☐1	☐2	☐3	td09
IS3G22I	i)	All <ethnic/racial> groups in the country have the same rights.	☐1	☐2	☐3	td05

Note: Variables names in the left side of each of the items are the original names present in public data files from ICCS 2016. In the right-hand side, we include the names td01-td09 to referred to the recoded responses analyzed in the present document. These responses were recoded so higher value expresses what is good for democracy. Items td06-td09 are reverse coded items, thus, for these items, higher values express what is bad for democracy.

The proposed items measure what a democratic system should look like (W. Schulz, Ainley, et al., 2018). to measure students conceptions of democracy (Judith Torney-Purta et al., 2006), and what is the meaning of democracy for students (Quaranta, 2019).

This collection of items have been present in the IEA Civic Education (CIVED) Study (J. Torney-Purta et al., 2001), and in the International Civic and Citizenship Education Study (ICCS) with different variations (W. Schulz et al., 2011; W. Schulz, Carstens, et al., 2018). These items present less research in comparison to other items and scales present in CIVED and ICCS studies (Knowles et al., 2018). We think this is the case because these responses present low common variance in a single trait model (ECV = .20) and throughout CIVED and ICCS studies IRT scores were not generated for these items (W.

Schulz et al., 2011; W. Schulz, Carstens, et al., 2018; Judith Torney-Purta et al., 2006). Thus, most of the previous research regarding these items exist using composite scores (Judith Torney-Purta et al., 2006) and descriptive results per item (W. Schulz, Ainley, et al., 2018; J. Torney-Purta et al., 2001; Torney-purta & Amadeo, 2004). However, two exemptions exist. One is the work from Husfeldt & Nikolova (2003), and more recently the work of Quaranta (2019). Husfeldt & Nikolova (2003). In the next section, we describe the approaches taken by these work to provide a sensible alternative regarding how to produce a standard for democracy conceptions measurement.

C.3.4.5.1 Previous modelling approaches

Husfeldt & Nikolova (2003) used data from CIVED 1999 and proposed three latent factors to modelled responses to a larger battery of items where most of the proposed items were included. These factors were "rights and opportunities", "limited government" and "threats to democracy". In the first factor, their work included items alluding to free speech, electing political leaders and protest against unjust laws. The second factor included items referring to free press, separation of church and state, and business having no restrictions. Finally, the third factor, for example, included items denoting nepotism, media control, coercion of justice by the government, among other indicators.

The work of Quaranta (2019) followed a different approach. The author used a person-centred analysis, to uncover interpretable patterns of responses between students. The author used a latent class analysis (Vermunt & Magidson, 2002), to reduce the observed responses to twelve items presented in ICCS 2009 of similar content to those items presented in ICCS 2016. In its research, the author found five different latent groups that distinguish students' responses regarding these items. These latent groups were named limited, free speech, minimalist, complex and uncritical. The limited class were students with low rates of 'strongly agree' responses to all items. Free speech class was characterized for a high rate of strongly agree only for the item referring to free speech ("Everyone should always have the right to express their opinions freely"). Minimalist class are students who strongly agree to items of free speech, that political rights should be respected for all people, that people should elect their political leaders and protest should never be violent. The complex class highly agree to items from the previous class, while also including a strong agreement to items referring no news media concentration, that people are able to criticize the government, protest against unfair laws and agree that differences in income between the rich and the poor should be small. Finally, the uncritical class are students how strongly agree to all the items, including positive and negative attributes for democratic systems. From out of these five latent classes, it seems the "complex" class seems the class closer to the intended interpretation of the SDG 4.7.4 subcategory of Democracy/democratic rule, democratic values/principles.

C.3.4.5.2 Item response theory modelling

In the present exercise, we explore the results from a unidimensional model including all items, and separate models for two different factors, more similar to the work of Husfeldt & Nikolova (2003). The generated scores using a partial credit model presented low reliability (EAP reliability = .57). This means respondents are too similar within this model, given the measurement error of the generated scores. We generated an IRT score including responses from items td01-td05, thus resembling factor 1 from Husfeldt & Nikolova (2003). Its resulting EAP reliability was also considerably low (EAP reliability

= .52) to provide trustworthy scores to generate standards. We proceed similarly with items td06-td09, resembling factor 3 ("threats to democracy") and we observed similar results regarding reliability (EAP reliability = .56). In summary, single latent trait models for all these items, and separate factors models, struggle to distinguish students' responses in reliable manner. In conclusion, these model approaches are discarded to represent democracy conceptions of students in a reliable manner.

C.3.4.5.3 Latent class analysis modelling

In the present report, we follow the approach of Quaranta (2019) and we fit a series of latent class analysis over the proposed items, including 1 to 10 latent classes. In particular, we specified a structurally homogenous model (Kankaraš & Vermunt, 2015). In practical terms, this model specifications searches for the same number of latent classes across countries, while keeping constant the types of expected response patterns across countries. Other models, such as the partially homogenous model specification (Kankaraš et al., 2011; Kankaraš & Vermunt, 2015) is a less interpretable model because it allows the pattern of responses to be different between countries while fixing only the amount of latent classes. Therefore, this later model is allowing differential item functioning for all items in all countries (Masyn, 2017). In practical terms, the structurally homogenous model specification allows the same interpretation of the pattern of responses across countries for each latent class. This property cannot be fulfilled with the partially homogenous model because it conforms to a country-specific model where all latent class can be different response patterns.

To estimate these models, we use Latent Gold 5.1 software (Vermunt & Magidson, 2013), including scaled survey weights (up to 1000), so each country contributed equally to the estimates (Gonzalez, 2012). For variance estimation, we use Taylor Series Linearization specifying primary sampling unit, and pseudo strata indicators (Asparouhov & Muthén, 2010; Stapleton, 2013). Before fitting the different latent class model, we recode the responses to each proposed item as dummy variables. Items td01-td05, where the response 1 is "Good for democracy" while the rest of the response categories were assigned a value of zero. Complementary, items td06-td09, were reverse coded, so a value of 1 was assigned to responses of "Bad for Democracy", while the rest of the response categories were coded as zero. We recode responses in this manner to avoid cells sparseness.

In the next table, the fit indexes of the ten fitted models are displayed.

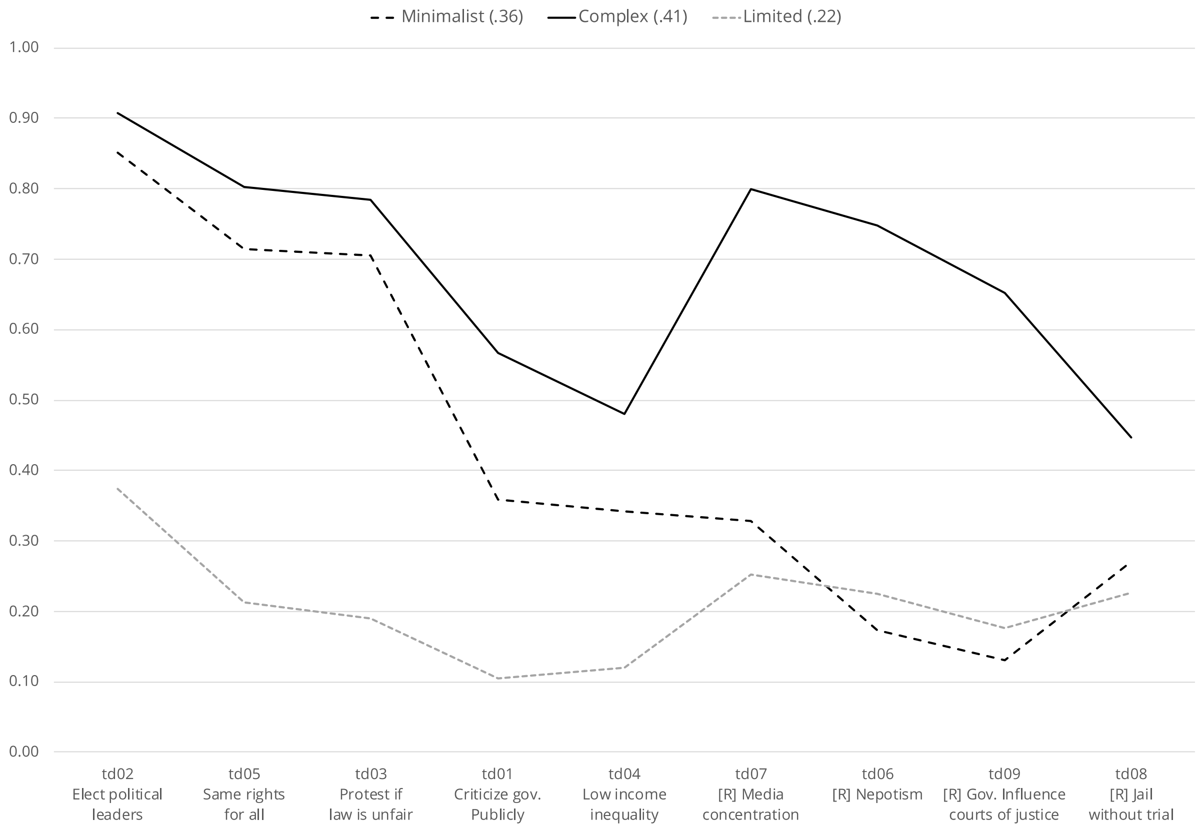
Table 25. Summary of fit indexes of the fitted latent class models

Classes	LL	BIC	Number of parameters	L^2	df	p-value	Classification error
1	-135137.91	270366.59	9	39679.30	23991	0.00	0.00
2	-127485.83	255395.25	42	24375.13	23958	0.03	0.12
3	**-125644.53**	**252045.51**	**75**	**20692.55**	**23925**	**1.00**	**0.18**
4	-124525.23	250139.74	108	18453.95	23892	1.00	0.23
5	-123904.30	249230.69	141	17212.07	23859	1.00	0.26
6	-123416.67	248588.28	174	16236.83	23826	1.00	0.29
7	-122943.33	247974.42	207	15290.14	23793	1.00	0.29
8	-122621.86	247664.31	240	14647.20	23760	1.00	0.31
9	-122300.49	247354.41	273	14004.47	23727	1.00	0.32
10	-122074.11	247234.48	306	13551.71	23694	1.00	0.33

Note: selected latent class model is highlighted in bold. LL = loglikelihood, BIC = Bayesian information criterion, L^2 = Likelihood ratio chi-square, df = degrees of freedom, p-value of the Likelihood ratio chi-square test. Classification error =

To decide regarding the most appropriate number of latent classes, we assess the models in terms of their fit to the observed data. The three latent class model fits the data well, presenting a good absolute fit to the observed data (L2= 20692.55, df = 23925, p = 1.00). These fit index results mean that the observed data can be generated by a fitted model of three latent classes (Masyn, 2013). This model presents a classification error of .18, which is the lowest classification error among all the fitted models with a satisfactory fit to the observed data (models with 3-10 latent classes). In the following figure, we present the response profile of the three latent class model.

Figure 23. Response patterns for What is good for democracy items from ICCS 2016



Because we are using a structurally homogenous across countries the response pattern or response profile is the same across countries. What is different, is the number of cases on each of these presented latent classes. To assigned names to the fitted latent classes, we used Quaranta (2019) latent group names. The minimalist class highly endorse the election of political leaders, the equal access to rights, and protest to unfair laws. However, is a less critical type, with less than 40% of endorsement for criticizing the government, and lower rates to threats for democracy, such as media concentration, nepotism, the influence of courts of justice by the government and jailing people without trial. This class represent 36% of students. The limited class, present low rates across all proposed items, thus failing to identify good and bad situations for democracy. In contrast, the students in the complex latent category, identifies as good for democracy electing political leaders, access to equal rights, and protesting if a law is unfair. Simultaneously, this class also identify as bad for democracy news media concentration, nepotism in the government, and the influence of government over the justice system.

We proposed to use the response pattern of the “Complex” latent class as the standard for the SDG 4.7.4 regarding the subcategory of Democratic principles. These are students who are more likely to identify situations that are deemed good for democracy, while at the same time are more likely to identify situations that are bad for democracy. In the next table, we include the expected percentages of these latent class realizations at the population level.

Table 26. Percentage of students meeting the SDG 4.7.4 Freedom (of expression, of speech, of press, of association/organisation)

Country or Region	**Complex**	**Minimalist**	**Limited**
Dominican Republic	0.03	0.81	0.16
Peru	0.09	0.75	0.16
Colombia	0.11	0.74	0.16
Mexico	0.11	0.63	0.26
Malta	0.25	0.46	0.28
Norway	0.30	0.51	0.19
Chile	0.34	0.41	0.25
Belgium (Flemish)	0.39	0.48	0.13
Latvia	0.40	0.10	0.50
Russian Federation	0.41	0.32	0.28
Lithuania	0.42	0.34	0.24
Bulgaria	0.42	0.34	0.23
Korea, Republic of	0.47	0.36	0.17
Italy	0.47	0.38	0.15
Sweden	0.51	0.35	0.14
Hong Kong SAR	0.51	0.25	0.24
Estonia	0.52	0.22	0.26
Netherlands	0.54	0.30	0.16
Slovenia	0.54	0.30	0.16
Denmark	0.56	0.24	0.20
Croatia	0.60	0.26	0.14
Finland	0.61	0.17	0.22
North Rhine-Westphalia	0.61	0.22	0.17
Chinese Taipei	0.77	0.06	0.18