Assessing pupils’ attitudes towards religious and worldview diversity – development and validation of a nuanced measurement instrument

Sabine Hermisson, Perman Gochyyev & Mark Wilson

To cite this article: Sabine Hermisson, Perman Gochyyev & Mark Wilson (2018): Assessing pupils’ attitudes towards religious and worldview diversity – development and validation of a nuanced measurement instrument, British Journal of Religious Education, DOI: 10.1080/01416200.2018.1556604

To link to this article: https://doi.org/10.1080/01416200.2018.1556604
Assessing pupils’ attitudes towards religious and worldview diversity – development and validation of a nuanced measurement instrument

Sabine Hermisson, Perman Gochyyev and Mark Wilson

Department of Religious Education, Protestant-Theological Faculty, University of Vienna, Vienna, Austria; Graduate School of Education, University of California, Berkeley, CA, USA; Berkeley Evaluation and Assessment Research (BEAR) Center, University of California, Berkeley, CA, USA

ABSTRACT
In this article, we outline the development and validation of an instrument for the assessment of attitudes towards religious diversity. The instrument uses four parallel scales, which evaluate attitudes towards Muslims, Christians, Jews and towards Non-religious people. Each scale is subdivided into eight parallel profile types and tests for the acceptance of six levels of social proximity. Drawing on a sample of 281 respondents (which include Muslims, Christians, Jews, Non-religious and Other) and employing the item response theory, we examine the reliability and validity of the instrument and present first results. We found that respondents discriminated greatly between the eight profile types and made clear distinctions with regard to the proximity of social relationships. Relative to these distinctions within each religion, differences between the four groups Muslims, Christians, Jews and Non-religious appeared minimal. These findings suggest that the results of previous research which indicate scepticism towards Islam despite general respect for religious diversity could be due to stereotyping.

KEYWORDS
Religious diversity; attitudes; multidimensional item response theory; anti-Semitism; Islamophobia; stereotype

1. Introduction
In the context of increasingly diverse societies, reliable data on young peoples’ attitudes towards religious plurality are one of many desiderata. How do young people perceive religions and worldviews that differ from their own? How do they feel about the increasing diversity of religions and worldviews? On the one hand, studies suggest that young people plead for respect for all religions (Yendell 2016; Francis, Penny, and McKenna 2016) and appreciate the acknowledgement of religious diversity in Religious Education and in school in general (Ipgrave 2016). On the other hand, adolescents appear to be sceptical of religious plurality, especially with regard to Muslims (e.g. Yendell 2016; Brockett, Village, and Francis 2009). These findings are in accordance with studies that report on prejudice towards Muslims (Yendell and Friedrichs 2012; Pickel 2013; Hafez and Schmidt 2015; Halm and Sauer 2017) and on Muslim pupils’ experiences of discrimination (e.g. Willems 2017).

The somewhat contradictory findings suggesting both respect and scepticism call for more fine-grained analyses. Who or what is it exactly that pupils find easy to tolerate or, conversely, hard to accept? Which (real or assumed) circumstances can cause respect and openness to turn into reserve and rejection? Our study aimed at developing a measurement instrument able to gauge such nuances. In the following, we describe its multi-cycled development, examine evidence for its reliability and validity, and report first results.
2. Theoretical framework and focus of the study

2.1. The research construct

The study of *attitudes* is a major field in social psychology. Attitudes are conceptualised as latent constructs composed of cognitive, affective and behavioural components (e.g. Eagly and Chaiken 1993; Zanna and Rempel 1988). We focus on the affective component without asking whether respondents are willing to accept diversity but rather how comfortable they are with it. This focus was chosen for two reasons. First, because emotion is an essential dimension of attitude, especially when it comes to religion. Second, focusing on the affective component allowed us to limit response bias due to social desirability. When questioned on a cognitive basis, many people tend to profess an attitude of acceptance towards diversity in an effort to conform to a socially desirable standard of behaviour. Our approach was to frame the conversation in such a way as to limit this effect. A similar strategy was employed by the Eurobarometer 2018 (European Commission 2018).

As to ‘religious and worldviews diversity’, there is no one universally accepted definition of the term (Bertram-Troost 2011). In this study, we decided to focus on attitudes towards Muslims, Christians, Jews and Non-religious people, because these are the four groups that are numerically strongest and most frequently covered by the media within our intended regions of study, Europe and North America. This allowed us to work on the assumption that young people have formed attitudes towards members of the groups in question. To avoid the continuous use of the cumbersome phrase ‘religious and worldview diversity’, ‘religious diversity’ is used hereafter in a broad sense that includes non-religious worldviews such as agnosticism and atheism.

More specifically, we concentrate on the acceptance of social proximity as the *content area*. Our main question is this: What profile types do pupils feel comfortable interacting with, as neighbours and classmates or more intimately as friends or family members? Acceptance of social proximity is a classic theme in the debate over migration (Wark and Galliher 2007), which made it possible for us to build on previous research.

2.2. Item response theory

Item response theory (IRT) – also known as latent trait theory – provides a rich array of psychometric tools to measure latent variables such as competences and attitudes. It has become state of the art in education research, particularly in the domains of math and science education (e.g. Wu and Adams 2006; Konnemann, Asshoff, and Hammann 2016). Its application to religious education research however has been limited so far (Benner et al. 2011; Riegel and Kindermann 2017). Unlike classical test theory – also known as ‘true score theory’ – in which analyses are executed on the entire test, IRT is a probabilistic approach that focuses on individual items. IRT has a number of advantages that have been driving its increased use. First of all, there is the fact that unbiased item estimates may be obtained from nonrepresentative samples and that, correspondingly, person estimates are independent of the specific sample of items administered to the respondents (Embretson 1996; Ayala 2013; Siddiq, Gochyyev, and Wilson 2017).

The simplest model within the IRT framework is the Rasch model (Rasch 1960). Based on the Rasch model, a number of additional models have been developed such as the Rating Scale Model (Andrich 1978) and the Partial Credit Model (PCM) (Masters 1982), which can both be applied to polytomous data. Furthermore, the unidimensional models have been expanded into a variety of multidimensional models to account for constructs consisting of several factors (Siddiq, Gochyyev, and Wilson 2017). Which of the models is used depends on the structure of the instrument (e.g. dichotomous or polytomous responses), on the research interest (e.g. measurement or explanation) and on a comparison of model fit after running analyses with different models under consideration (Boeck and Wilson 2004).
2.3. Research question

The objective of our study was to establish a construct for attitudes towards religious diversity and to develop a corresponding measurement instrument. To this end, we addressed the following questions in particular:

1. What evidence can be presented to support the internal validity (e.g. reliability, item fit, internal structure) and the external validity of the instrument (e.g. correlation to background variables)?
2. Does the measurement instrument produce reliable and valid results for adolescents between 14 and 15 years of age, and between 16 and 18 years of age?
3. Is ‘acceptance of religious diversity’ a unidimensional or multidimensional construct? i.e. do people differ in their degree of acceptance towards Muslims, Christians, Jews and the Non-religious?
4. How are attitudes towards each of the four groups (Muslims, Christians, Jews and the Non-religious) related to background variables such as religious affiliation?

3. Method

3.1. Multi-cyclical development of the measurement instrument

We employed the Construct Modeling approach suggested by Wilson (2005). Construct and measurement instrument were developed in multiple cycles.

In a first cycle, we conducted qualitative interviews with pupils (age 14–15, n = 2) which included items from the survey by Francis, Penny, and McKenna (2016). The interviews indicated that the precision of the items would be decisive for the reliability and validity of the instrument. This is reflected in the following interview passage with a 14-year-old girl:

I (interviewer): ‘All religious groups must have equal rights.’ What do you think?
R (respondent): I’m not sure. If ‘religious groups’ means religions like Islam or Christianity or so, I would say ‘strongly agree’. But if it also means like the Flying Spaghetti Monster, then I’d rather choose ‘not certain’. […]
I: ‘I would be happy about a close relative marrying someone from a different faith.’ How do you feel about this?
R: If my sister wants to marry a Muslim and he is open, I’d choose ‘strongly agree’. But if he wants her to wear a head scarf, I don’t know. I wouldn’t say ‘I’m happy’. I’d say ‘agree’.

The interview passage illustrates: The interviewee’s responses were to a considerable degree determined by her interpretation of the items and not simply by her attitudes towards religious diversity. The phrasing of the item (‘all religious groups’, ‘a Muslim’) leads her to consider a number of different interpretations: world religions versus the flying spaghetti monster, open-minded versus restrictive Muslim.

In a second cycle, we developed a construct map, items and a scoring guide using expert panelling and cognitive think-aloud interviews (Hermisson 2017). We were able to draw on individual items from the study by Francis, Penny, and McKenna (2016, see the interview above) and from the survey by Halm and Sauer (2017), which included an item on the acceptance of Muslims as neighbours. The reported findings from the qualitative interviews prompted us to build a double differentiation into the structure of the items: (a) Instead of testing for the acceptance of ‘a Muslim’, a spectrum of 12 specific profiles was developed ranging from secular to fundamentalist with parallel profiles for Christians, Jews and the Non-religious. (b) In addition to ‘neighbour’, we included additional levels of social proximity such as friendships and family relationships to test where respondents draw the line.
In a third cycle, we conducted a pilot study \((n = 114)\) to test the construct map and the measurement instrument with regard to its calibration, reliability and validity (Hermisson 2017). This included two think-alouds (age 14–15, \(n = 2\)) and four exit interviews (age 14–15, \(n = 2\), age \(\geq 25\), \(n = 2\)).

In a fourth cycle, we revised the construct map on the basis of the outcomes of the pilot study, eliminated redundant items and items with poor fit. Four additional items were developed to close a calibration gap.

Finally, using a spectrum of \(4 \times 8\) items, we carried out the present study \((n = 281)\), including think-alouds (age: 14–15, \(n = 3\)) and exit interviews as before (age: 14–15, \(n = 3\)).

### 3.2. Item design: 32 profiles and 6 levels of social proximity

The item design process resulted in: (a) a spectrum of \(4 \times 8\) profile types and (b) six levels of social proximity.

(a) The spectrum of \(4 \times 8\) profile types is presented in Table 1

| Level | Muslims                                                                 | Christians                                                                 | Jews                                                                 | Non-religious people                                                                 |
|-------|-------------------------------------------------------------------------|----------------------------------------------------------------------------|----------------------------------------------------------------------|-----------------------------------------------------------------------------------|
| 8     | A Muslim who, on the basis of her/his faith, is against religious freedom | A Christian who, on the basis of her/his faith, is against religious freedom | A Jew who, on the basis of her/his faith, is against religious freedom | A non-religious person who, on the basis of her/his worldview, is against religious freedom |
| 7     | A Muslim who tries to convert me to (his/her version of) Islam          | A Christian who tries to convert me to (his/her version of) Christianity   | A Jew who tries to convert me to (his/her version of) Judaism        | A non-religious person who tries to convert me to (his/her version of) Atheism       |
| 6     | A Muslim who repeatedly argues that any other faith or worldview is wrong | A Christian who repeatedly argues that any other faith or worldview is wrong | A Jew who repeatedly argues that any other faith or worldview is wrong | A non-religious person who repeatedly argues that any other worldview or faith is wrong |
| 5     | A Muslim who strictly follows Muslim teachings in ethical and political questions (e.g. human rights, war and peace) | A Christian who strictly follows Christian teachings in ethical and political questions (e.g. human rights, war and peace) | A Jew who strictly follows Jewish teachings in ethical and political questions (e.g. human rights, war and peace) | A non-religious person who strictly follows an Atheist worldview in ethical and political questions (e.g. human rights, war and peace) |
| 4     | A Muslim who wears religious symbols or clothing in public (e.g. the female head scarf called hijab) | A Christian who wears religious symbols or clothing in public (e.g. a T-shirt with the message ‘Jesus saves’) | A Jew who wears religious symbols or clothing in public (e.g. the head covering for males called kippah) | A non-religious person who wears symbols or clothing stating her/his worldview (e.g. a T-shirt with the message ‘Atheist’) |
| 3     | A Muslim who talks about his/her faith in public                        | A Christian who talks about his/her faith in public                        | A Jew who talks about his/her faith in public                        | A non-religious person who talks about his/her non-religious worldview in public |
| 2     | A Muslim who practices his/her religion only in private                | A Christian who practices his/her religion only in private                | A Jew who practices his/her religion only in private                | A non-religious person who keeps his/her non-religious worldview private            |
| 1     | A Muslim who is not practicing her/his religion (e.g. doesn’t pray, doesn’t go to the mosque) | A Christian who is not practicing her/his religion (e.g. doesn’t pray, doesn’t go to church) | A Jew who is not practicing her/his religion (e.g. doesn’t pray, doesn’t go to the synagogue) | A non-religious person who doesn’t bother about belief or non-belief                |
relational clothing, Level 4). The items were developed, tested and revised with a view to the optimal calibration of the instrument. This means that the spectrum was designed to cover the full range of item difficulty. We hypothesised that item difficulty would increase from each level (1–8) to the next for each of the four scales.

(b) Regarding the levels of social proximity, we drew on the Bogardus Social Distance Scale (Wark and Galliher 2007), a well-established scale in the social sciences. Since the Bogardus Scale stems from an era of patriarchal family structures (it was created in 1924) and was designed for adult males, we adapted it for postmodern family structures, including patchwork families, and reformulated it for young people as respondents. Note that the original Bogardus Scale as well as our adaptation implies a Guttman scaling approach (Guttman 1944, 1950; Wilson and Gochyyev 2013). Guttman’s scalogram is based on the assumption that a respondent who endorses a more extreme statement will also endorse all less extreme ones (e.g. a respondent who feels comfortable accepting a ‘Muslim who strictly follows Islamic teaching’ (m5) as a ‘close friend’, will also feel comfortable accepting him/her on a less intimate level (e.g. as a neighbour). Designing an instrument that conforms to such a strict framework has advantages for the statistical analysis, while requiring careful item and item-category specifications. This way the instrument allows for polytomous scoring of the items rather than five successive dichotomous scorings for each individual level of social proximity. In the Rasch modelling approach used in Construct Modelling (Wilson 2005), this structure is softened to a probability.

The empirical data from the pilot study (n = 114) indicated that the Guttman scale applied to four of the six levels of social proximity, while the other two levels (family relationship vs. close friendship) did not display a clear hierarchical order. Therefore, we revised the levels and the scoring as to better conform to the Guttman structure (see Table 2). That this in fact worked was confirmed by the present study.

3.3. The questionnaire

The questionnaire contained 32 items employing the 4 × 8 profile types of Muslims, Christians, Jews and Non-religious specified in Table 2. For each individual profile, respondents were

| Level | Hypothesised structure                                      | Revised structure                                      | Outcome space           | Score |
|-------|-------------------------------------------------------------|--------------------------------------------------------|-------------------------|-------|
| 6     | Acceptance of religious diversity on the level of partner relationship | Acceptance of religious diversity on the level of partner relationship | (future) Spouse or partner | 5     |
| 5     | Acceptance of religious diversity on the level of family relationship | Acceptance of religious diversity on both levels: family relationship and close friendship | Immediate family member and close friend | 4     |
| 4     | Acceptance of religious diversity on the level of close friendship | Acceptance of religious diversity on the level of family relationship or close friendship | Immediate family member or close friend | 3     |
| 3     | Acceptance of religious diversity on the level of everyday social contacts | Acceptance of religious diversity on the level of everyday social contact | Neighbour or classmate | 2     |
| 2     | Acceptance of religious diversity on the level of fellow citizenship | Acceptance of religious diversity on the level of fellow citizenship | Fellow citizen | 1     |
| 1     | No acceptance of religious diversity on any level of social proximity | No acceptance of religious diversity on any level of social proximity | None of the above | 0     |
asked to indicate whether they felt comfortable having the person as: a fellow citizen, neighbour or classmate, close friend, immediate family member, (future) spouse or partner, or none of the above. The formulation ‘I feel comfortable’ was used to gauge the affective dimension of the respondents’ attitudes as discussed in Section 2.1. Table 3 presents an excerpt of the questionnaire.

Both an online version and a paper-and-pencil version of the questionnaire were developed. Note that, when a respondent chose more than one category, the highest level of proximity was given the resulting score.

3.4. Independent variables

The background questionnaire consisted of five additional questions regarding age, gender, religious affiliation (Muslim, Christian, Jew, Buddhist, Hindu, Non-religious or Other), personal importance of one’s own religion/non-religious worldviews (ranging from ‘not important’ to ‘very important’) and frequency of contacts with Muslims, Christians, Jews and Non-religious, respectively (ranging from ‘a lot’ to ‘not at all’).

3.5. Sample

The data were gathered in the San Francisco Bay area in a Presbyterian Christian congregation, a Reform Jewish congregation, a Muslim community and a secular Sunday assembly. The data collection took place between May and July 2017. The sample consisted of 281 respondents. In terms of religious affiliation, respondents identified as Muslims (n = 56), Christian (n = 139), Jews (n = 29), Non-religious (n = 47), Buddhists (n = 2), Hindu (n = 1) and ‘Other’ (e.g. ‘Post-Christian’, ‘Christian and Jew’, ‘Christian and Non-religious’) (n = 7). With regard to gender, 175 respondents identified as female, 103 as male and 3 as other or not specified. As to age, respondents were: ≤13 years (n = 2), 14–15 years (n = 10), 16–18 years (n = 17), 19–24 years (n = 20), 25–34 years (n = 33), 35–49 years (n = 62), ≥50 (n = 135) and unidentified (n = 2).

3.6. Multidimensional and polytomous item response modelling

We accounted for the polytomous nature of the responses by applying the PCM (Masters 1982). To model the four correlated profile groups, we used the multidimensional version of the PCM. The Multidimensional Random Coefficients Multinomial Logit Model (MRCML) was proposed by Adams, Wilson, & Wang (1997) as the variant of the multidimensional IRT model in the Rasch modelling framework. MPlus 8 (Muthén and Muthén 1998–2017) and ConQuest 3 (Adams, Wu, and Wilson 2012) were used for the estimation of all of the measurement models presented in the paper.

### Table 3. Excerpt from the questionnaire.

| Each set of items below presents a different profile of a Christian. For each profile, please check all the answers that fit for you. |
| You can check 0–5 boxes for each profile. Or you can check ‘none of the above’. |
| (1) I feel comfortable having a Christian who is not practicing his/her religion (e.g. doesn’t pray, doesn’t go to church) as a |
| ☐ fellow citizen  |
| ☐ neighbour or classmate  |
| ☐ close friend  |
| ☐ immediate family member (e.g. stepsister, brother-in-law)  |
| ☐ (future) spouse or partner  |
| ☐ none of the above. |
3.7. Latent regression

We compared attitudes towards the four groups (Muslims, Christian, etc.) across background variables (e.g. gender, age, etc.) and used the so-called latent regression or item response model with manifest predictors (Mislevy 1987; Verhelst and Eggen 1989; Zwinderman 1991, 1997) to account for measurement errors in the latent variables when investigating the relationship to the background variables. We thus regressed the actual latent variable itself within the model. This approach is a better alternative to regressing predictions (e.g. EAP scores) of the four latent variables on background variables (Adams, Wilson, & Wang 1997).

4. Results

4.1. Evidence of internal validity

We applied the unidimensional PCM and the four-dimensional PCM to examine model fit and to test for the dimensionality of the instrument. As indicated in Table 4 below, the difference in deviances between the two models is 348 (16,055–15,707) with 12 (14–2) degrees of freedom, which is statistically significant at 0.005 (based on a conservative test described by Rabe-Hesketh and Skrondal 2005). The difference in deviance implies that the four-dimensional PCM fits significantly better than the unidimensional PCM. This is also supported by the Akaike Information Criterion (AIC).

The number of parameters does not include the item parameters which were fixed ('anchored') to the values obtained from the dimensional alignment procedures (Schwartz, Ayers, and Wilson 2017). The parameters in the unidimensional model are the mean and variance for the single dimension. The parameters in the four-dimensional model are four means, four variances and six covariances between the four dimensions. The correlation among the four dimensions (i.e. attitudes towards Muslims, Christians, Jews, and Non-religious) is highest for Christians and Jews (estimated at 0.979) and lowest for attitudes towards Muslims and Non-religious (estimated at 0.685) (see Appendix – Table A1).

In order to provide evidence for the internal validity of the instrument, we investigated the internal consistency of the instrument (see Table 5). For the four-dimensional model, our findings show dimension-specific reliabilities of 0.95, 0.93, 0.98 and 0.95 for attitudes towards Muslims, Christians, Jews and Non-religious, respectively, and an overall reliability of 0.91. The type of reliability reported here is the Expected A Posteriori (EAP) reliability (Adams, Wu, and Wilson 2012) which is commonly used in the IRT framework. EAP and Cronbach’s alpha of 0.91 (Cronbach 1951) indicate a high reliability of the instrument, both for the overall instrument and for each of the four individual scales.

Table 4. Model summaries for the unidimensional and multidimensional PCM for 32 items.

| Model               | Deviance | AIC   | Number of parameters |
|---------------------|----------|-------|----------------------|
| Unidimensional PCM  | 16,055   | 16,059| 2                    |
| Four-dimensional PCM| 15,707   | 15,735| 14                   |

Table 5. Item analysis results for the unidimensional analysis of DI (column 1) and the multidimensional analysis of DI (column 2).

| Statistics                  | Unidim. model | Multidim. model |
|-----------------------------|---------------|-----------------|
| Sample size                 | 281           | 281             |
| Number of items in the calibration | 32            | 32              |
| Missing data                | 1.5%          | 1.5%            |
| Model                       | PCM           | PCM             |
| Weighted fit MNSQ >1.33, T sig at 0.05. (Item Parms) | Two items | None |
| Weighted fit MNSQ >1.33, T sig at 0.05. (Step Parms) | Three steps | 1 step |
| Reliability estimates:      |               |                 |
| EAP reliability             | 0.92          |                 |
| Attitude towards Muslims    | 0.95          |                 |
| Attitude towards Christians | 0.93          |                 |
| Attitude towards Jews       | 0.98          |                 |
| Attitude towards Non-religious | 0.95       |                 |
| Cronbach’s alpha            | 0.91          | 0.91            |
The item fit statistics suggest that all the items fit reasonably well within the usual range. To check whether the items are well aligned with the multidimensional Rasch model, the weighted mean square fit statistics are estimated for each item. Note that item and item-step parameters were fixed for the dimensional alignment of the four scales. More specifically, we used the delta-dimensional alignment (DDA) technique (Schwartz 2012), since each of the four factors has its own scale in the four-dimensional Rasch model. For the multidimensional model, we found that all of the 32 items have less than 4/3 (1.33) weighted mean square fit used as acceptable upper bound (Adams and Khoo 1996) with only one of the 128 item steps falling outside (see Table 5).

Finally, the empirical results are in accordance with the hypothesised structure of the instrument. We had hypothesised the difficulty of the profile types to increase from each level (1–8) to the next for each of the four scales (Muslims, Christians, Jews, Non-religious). Overall, the empirical results confirm this hypothesis. The values presented in Table 6 indicate the estimated item difficulties for each of the 32 items. Low values indicate easy items. For each of the scales, item difficulty increases from level to level. Consider, e.g. the Muslims scale: The estimated difficulty of m1 (= Muslim item on Level 1) is lower than m2, m2 is easier than m3 etc. Only 5 out of the 32 items breach the hypothesised structure (j3, n2, m7, c7, j7).

The fact that the empirical results are, for the most part, in accordance with the hypothesised internal structure of the instrument can be considered further evidence of its validity.

The visual representation presented in Figure 1 provides the cumulative category characteristic curves for the response choices (increasing levels of social proximity). The curves are averaged across all items and show the probability of selecting a particular category or higher (i.e. ‘[fellow citizen] OR higher’, ‘[neighbour or classmate] OR higher’ and so on). Note that selecting ‘[none of the above] OR higher’ implies any category, hence the curve for this is not shown (the probability is one). Regarding the six levels of social proximity, we found that the thresholds from one level to the next are fairly equally distributed for Levels 1–5 (from ‘none of the above’ to ‘close friend and family member’). However, the threshold to Level 6 (to (future) spouse or partner) appears to be much harder to cross, as illustrated by the curve being further apart, and has less discriminatory power as illustrated by the flatness of the curve.

This anomaly could be a valid result in accordance with the construct as intimate personal relationships accentuate differences. However, exit interviews indicated that other reasons could play a role as

Table 6. Estimates of the item difficulties and standard errors (in logits) after dimensional alignment (for valid comparisons across dimensions).

| Level | Muslims                  | Christians                | Jews                    | Non-religious             |
|-------|--------------------------|---------------------------|-------------------------|--------------------------|
| 8     | 1.717 (0.088)            | 1.619 (0.083)             | 1.499 (0.080)           | 1.621 (0.079)            |
|       | M against religious freedom | C against religious freedom | J against religious freedom | N against religious freedom |
| 7     | 0.719 (0.061)            | 0.697 (0.061)             | 0.744 (0.061)           | 0.648 (0.058)            |
|       | M trying to convert      | C trying to convert       | J trying to convert     | N trying to convert      |
| 6     | 0.790 (0.062)            | 0.717 (0.060)             | 0.874 (0.063)           | 0.592 (0.057)            |
|       | M arguing that any other faith or worldview is wrong | C arguing that any other faith or worldview is wrong | J arguing that any other faith or worldview is wrong | N arguing that any other faith or worldview is wrong |
| 5     | –0.919 (0.062)           | –0.957 (0.062)            | –0.934 (0.062)          | –0.661 (0.060)           |
|       | M strictly following     | C strictly following      | J strictly following    | N strictly following     |
|       | Muslim teachings         | Christian teachings       | Jewish teachings        | Atheist worldview        |
| 4     | –1.534 (0.081)           | –0.979 (0.062)            | –1.588 (0.079)          | –0.811 (0.065)           |
|       | M wearing religious symbols or clothing (hijab) | C wearing religious symbols or clothing (T-shirt 'Jesus saves') | J wearing religious symbols or clothing (kippah) | N wearing symbols or clothing stating her/his worldview |
|       | (T-shirt 'Jesus saves')  | (T-shirt 'Jesus saves')   | (kippah)                | (T-shirt 'Atheist')      |
| 3     | –1.766 (0.084)           | –1.647 (0.077)            | –1.817 (0.078)          | –1.569 (0.080)           |
|       | M talking about faith    | C talking about faith      | J talking about faith    | N talking about worldview |
| 2     | –1.860 (0.092)           | –1.813 (0.090)            | –1.809 (0.089)          | –1.914 (0.093)           |
|       | M practicing only in private | C practicing only in private | J practicing only in private | N keeping worldview private |
| 1     | –1.983 (0.096)           | –1.823 (0.090)            | –2.192 (0.097)          | –1.656 (0.086)           |
|       | M not practicing         | C not practicing           | J not practicing         | N not practicing          |
well. One respondent in particular expressed that she preferred to remain celibate. Therefore, in order not to potentially compromise the validity of our instrument, we collapsed the two levels ‘(future) spouse or partner’ and ‘close friend and family member’ for the analysis reported below. The spouse/partner category yields relevant insights that are worth exploring elsewhere nonetheless.

4.2. External validity evidence

The validity of the measurement instrument is also supported by correlation to the background variable ‘religious affiliation’. We had hypothesised that respondents would favour relationships with others from the same religious background. As discussed in detail in Section 4.5, this was confirmed in the analysis of the data. Therefore, the background variable ‘religious affiliation’ is external evidence for the validity of the instrument.

4.3. Does the instrument work for adolescents?

Since the results reported so far are all based on a sample consisting of respondents across all ages, the question of how well the measurement instrument works for adolescents is yet to be examined. Both evidences based on the response process as well as on the correlation with the background variable ‘age’ suggest that the reliability and validity of the instrument are high for adolescents aged 14 years and older. First of all, the instrument was geared to include this age group and adolescents were involved in the formative phase of the instrument (see above Section 3.1). Second, all of the five think-alouds and two of the five exit interviews were conducted with adolescents (all age 14–15) and indicated that young people are well capable of understanding the structure and content of the questionnaire and the individual items. This is confirmed, finally, by person-fit statistics, which assess the validity of the instrument at the individual level and are based on the consistency of an individual’s item response pattern (Embretson and Reise 2000, 238). Checking for person fit allows us to identify individuals with aberrant response patterns that are e.g. due to misunderstandings of the questionnaire or individual items. Person-fit statistics (Table 7) indicated that all age groups, including adolescents age 14–15 and 16–18 years, exhibited fairly consistent response patterns and are well within the expected range of less than 4/3 (1.33).
4.4. Capability of the measurement instrument

The development process resulted in a dense and complex measurement instrument. For any given sample, it allows answers to a large variety of specific research questions surrounding religious diversity. These include the following:

4.4.1. Individual items of a scale
What do respondents perceive as ‘easy’, what as ‘difficult’ to accept? E.g. is it easier to accept a Christian who practices only in private than a Christian who talks about his/her faith in public?

4.4.2. Classical item discrimination
Which items are most sensitive for measuring respondents’ attitudes?

4.4.3. Parallel items of the four scales
E.g. what is easier to accept: a hijab, a kippah, a ‘Jesus saves’-T-shirt or an ‘Atheist’-T-shirt?, a Christian or an Atheist eager to convert?

4.4.4. Difficulty of the four scales
How comfortable are respondents with Muslims, Christians, Jews and the Non-religious in general? Is there a particular group that is discriminated against?

4.4.5. External variable religious affiliation
How do respondents with affiliations to individual religions or worldviews relate to religious diversity and to each one of the four groups (Muslims, Christians, Jews and Non-religious)? Is there a tendency for Christian anti-Semitism, Muslim anti-Semitism or Jewish Islamophobia?

4.4.6. External variables of age, gender and frequency of contact
Are young respondents more comfortable with religious diversity than older respondents? Are there gender-specific differences in the attitudes towards Muslims, Christians, Jews and the Non-religious or in the acceptance of single items (e.g. hijab)? Is personal contact correlated with a higher level of acceptance, as Allport’s contact hypothesis (1954) postulates?

Due to limited space, we will only demonstrate the use of the instrument by presenting a few of our findings. Detailed results will be given elsewhere. We focus on three issues: (a) on item and scale difficulty, (b) on the external variable ‘religious affiliation’ and its effect on acceptance and (c) on the religious clothing question.

4.5. Selected findings

The Wright map in Figure 2 provides a visual representation of many of our main findings. The left side of the Wright map shows the respondents’ estimated mean level of acceptance for each of the four scales. Each ‘X’ represents 2.8 respondents. The most accepting respondents are placed at the

| Table 7. Mean person fit per age group. |
|-----------------------------|-----------------|
| Age 14–15 | 0.792 | 0.558 |
| Age 16–18 | 0.790 | 0.502 |
| Age 19–24 | 0.722 | 0.592 |
| Age 25–34 | 0.712 | 0.859 |
| Age 35–49 | 0.601 | 0.509 |
| Age ≥50  | 0.594 | 0.529 |
| Total    | 0.639 | 0.578 |
Figure 2. Wright map for the four scales.
top; the least accepting respondents are placed at the bottom. The right side of the map indicates the estimated mean difficulty for each of the 32 items.

The map indicates the following:

4.5.1. Not ‘the’ Muslim or ‘the’ Christian but considerable diversification
Respondents varied greatly in their acceptance of the eight profile levels, ranging roughly from –2 to +4 logits. While most respondents felt comfortable having close relationships with secular Muslims, Christians, etc. (m2, c2, j2, n2), they preferred not to interact with Muslims, Christians, etc. who are against religious freedom (m8, c8, j8, n8).

4.5.2. Tipping point
The large gap between the items of Level 5 (‘strictly following Muslim/Christian etc. teaching’) and Level 6 (‘arguing that any other faith or worldview is wrong’) indicate that respect for other faiths and worldviews or a lack thereof marks a bifurcation. Acceptance diminished greatly when other belief systems were demeaned (m6, c6, j6, n6). This gap persisted even though we had aimed for a balanced calibration and revised the items accordingly after the pilot study.

4.5.3. Equal difficulties of the four scales
We had anticipated divergent difficulties of the four scales (attitudes towards Muslim, Christian etc.). In the Wright map, this would have resulted in a lower distribution of the ‘XXX’s, say for the Muslim scale. However, contrary to our expectations, none of the four groups were perceived as significantly harder to accept by the respondents as a whole. This lack of discrimination is elaborated on in greater detail by the mean estimates presented in Table 8. The values indicate that the maximum difference in difficulty between the individual scales is less than 0.2 logits. None of these are statistically significant.

4.5.4. Similar difficulties of parallel items
This lack of discrimination applies equally to the level of individual items represented on the right side of the Wright map. See, e.g. the four most difficult items on the top: The estimated conditional difficulty of accepting someone who is against religious freedom does not differ greatly whether the person is Muslim (m8), Christian (c8), Jew (j8) or Non-religious (n8).

4.5.5. Hijab easier to accept than ‘Jesus saves’- or ‘Atheist’-T-shirt
While respondents found most parallel items similarly easy or difficult to accept, there are exceptions. This is particularly true for the clothing item which is the least parallel of the 8-item levels for substantive reasons. As the Wright map and Table 7 indicate, respondents felt more comfortable accepting a kippah (j4, −1.588 logits) or a hijab (m4, −1.534 logits) than a ‘Jesus saves’-T-shirt (c4, −0.979 logits) or an ‘Atheist’-T-shirt (n4, −0.811). Thus, even though all four pieces of clothing display personal beliefs, the two rooted in traditional religious practice enjoy higher degrees of acceptance. This is not only due to favourable ratings received from the own group, but also from the other groups.

| Table 8. Scale difficulties in logits, indicating no significant differences in the difficulties between the four scales (the lower the mean estimated value, the easier the scale). |
| Scale     | Mean   | SD    |
|-----------|--------|-------|
| Muslim    | −0.605 | 1.458 |
| Christian | −0.523 | 1.343 |
| Jewish    | −0.653 | 1.460 |
| Non-religious | −0.469 | 1.287 |
| Total     | −0.562 | 1.322 |
Additional findings relate to the background variable of ‘religious affiliation’. Table 8 below shows the difference in attitudes towards each of the four groups Muslims, Christians, Jews and Non-religious regressed on the respondents’ self-identification. Note that we ran a latent regression model, in which we controlled for age and gender but that those results are not indicative of significant differences (see Appendix – Table A2). We therefore report the values from Table 9, because here the estimates of the reference category are more directly interpretable as discussed in the comments below. The higher the value of a given group of respondents for any give scale, the more accepting these respondents are towards members of this religion.

The values indicate:

4.5.6. Preference for one’s own tribe
As discussed in Section 4.2, respondents accepted closer social relations with people from their own religious background – even when intimate relationships were disregarded. However, the differences that respondents made between each of the four groups are not very large in some cases (e.g. a maximum of 0.32 logits for Christian respondents, Row 2) whereas others are more pronounced (e.g. a maximum of 1.12 logits for Non-religious respondents, Row 4). Only Jewish respondents slightly favoured Non-religious profiles over Jewish ones. This is most likely due to the fact that some of the Jewish respondents from a Reformed Jewish congregation identify as cultural Jews and Atheists and proved to be fairly reserved towards the most extreme Jewish profiles.

4.5.7. Equality for all
Respondents as a whole were fairly balanced in their acceptance of religions and worldviews other than their own apart from the own group and, by and large, did not discriminate between Muslim, Christian, Jewish or Non-religious otherness. Consequently, no group was at a general disadvantage. We found neither Muslim nor Christian anti-Semitic tendencies nor Christian or Jewish anti-Muslim tendencies. Muslim respondents, while fairly reserved towards the three other groups, made no significant difference in their acceptance of Christians (−0.47) versus Jews (−0.56). Christian respondents differed only slightly in their acceptance of Jews (0.37) versus of Christians (0.45) and were more open to social interaction with Jews from the whole spectrum of Jewish profiles (0.37) than Jewish respondents (0.11). Similarly, we found neither Christian nor Jewish anti-Muslim tendencies: Christian respondents were found to be equally accepting of Muslims (0.14) as of the Non-religious (0.13). Jewish respondents did not differ significantly in their acceptance of Muslims (−0.37) or of Christians (−0.36). Non-religious respondents differed most pronouncedly in their acceptance of their own group (0.82) versus the other groups (−0.30, −0.27, −0.08). In contrast, the distinctions they drew between these three religious groups are minor.

Table 9. Mean estimates of attitudes and standard errors (in logits) for five religion categories in four dimensions.

|                    | Attd. towards Muslims | Attd. towards Christians | Attd. towards Jews | Attd. towards Non-religious | Δ       |
|--------------------|-----------------------|--------------------------|-------------------|-----------------------------|---------|
| Muslim respondents | 0.34 (0.14)           | −0.47 (0.13)             | −0.56 (0.15)      | −0.65 (0.14)                | 0.99    |
| (n = 56)           |                       |                          |                   |                             |         |
| Christian respondents | 0.14 (0.08)       | 0.45 (0.07)              | 0.37 (0.09)       | 0.13 (0.08)                | 0.32    |
| (n = 139)          |                       |                          |                   |                             |         |
| Jewish respondents | −0.37 (0.14)         | −0.36 (0.13)             | 0.11 (0.15)       | 0.18 (0.13)                | 0.54    |
| (n = 29)           |                       |                          |                   |                             |         |
| Non-religious      | −0.30 (0.14)         | −0.27 (0.13)             | −0.08 (0.15)      | 0.82 (0.12)                | 1.12    |
| respondents        |                       |                          |                   |                             |         |
| (n = 47)           |                       |                          |                   |                             |         |
| Other respondents  | −0.35 (0.13)         | −0.23 (0.11)             | −0.21 (0.13)      | −0.07 (0.09)               | 0.28    |
| (n = 10)           |                       |                          |                   |                             |         |
4.5.8. Greatest divide between religious and non-religious groups
The estimates presented in Table A1 also suggest that greatest divide is not between the different world religions but rather between religious and non-religious groups. This conclusion is also supported by findings on individual items.

5. Conclusion
We have developed a nuanced and versatile instrument, using item response theory as the psychometric approach, to measure attitudes towards religious and worldview diversity. In comparison to previous research, the instrument allows for a more fine-grained analysis. This was accomplished by testing for the acceptance of a spectrum of eight specific social relationship profiles of Muslims with parallel profiles for Christians, Jews and Non-religious across six levels of social proximity.

Internal evidence (internal consistency, item fit statistics and the empirical confirmation of the hypothesised structure) as well as external evidence (the correlation to the background variable ‘religious affiliation’) provide evidence for the reliability and validity of the developed instrument. This also holds true for adolescent respondents (from the age of 14 years).

We found that the four-dimensional PCM is a statistically significant better fit than its unidimensional version, with acceptance of the Non-religious to acceptance of Muslims having the lowest correlations (at 0.69), indicating that ‘acceptance of religious diversity’ is a multidimensional construct. In all four groups, respondents accepted closer social proximity with people from their own religious background. However, respondents did not discriminate significantly among the three groups other than their own. That is, the differences discovered between the four groups Muslims, Christians, Jews and Non-religious were minor. No religious group was at a general disadvantage – even with regard to the often-disputed hijab.

In contrast, respondents discriminated strongly between the eight profile types of each scale and made clear distinctions with regard to social relationships. They specifically indicated noticeable reserve towards profile types that lack respect for other faiths or worldviews. Our data thus suggests that the distinction between the eight profile types is highly relevant.

Previous research indicated that young people, even though they profess respect for all religions (Yendell 2016; Francis, Penny, and McKenna 2016), are sceptical about religious plurality, especially with regard to Muslims (e.g. Yendell 2016; Brockett, Village, and Francis 2009). The fact that we did not discover comparable levels of discrimination towards Muslims could be due to the nature of the sample, since the data was collected in a region where diversity is valued. Comparing these results to a European sample in the future may prove interesting. However, the data suggest a more interesting hypothesis. The results of previous research may be explained through correlation of a particular religion (e.g. Islam) with a distinct profile type (e.g. a Muslim strictly following Muslim teaching, m5) due to stereotyping. Given the considerable differences that respondents made between the eight profile types within a particular religion, even a weak correlation due to stereotyping can have significant effect. Testing this interpretation is a worthwhile avenue for further research, for which the instrument provides a necessary tool.

Disclosure statement
No potential conflict of interest was reported by the authors.

Notes on contributors
Sabine Hermisson is an assistant professor at the Institute for Religious Education, Protestant Theological Faculty, University of Vienna, Austria. Her research interests include attitude research, religious diversity, theology and natural science, qualitative and quantitative methods. In 2016/2017, she was a visiting scholar at the Pacific School of Religion and the Center for Theology and Natural Science, Berkeley.
Perman Gochyeev is a behavioural statistician and a project manager at the Graduate School of Education University of California, Berkeley, and at the Berkeley Evaluation and Assessment Research (BEAR) Center. His research areas include psychometric and multilevel modelling, and evaluation. Prior joining BEAR Center, he was the Manager of Research and Analytics at Medallia’s CX Strategy Research group and a research associate at WestEd.

Mark Wilson is a professor at the Graduate School of Education University of California, Berkeley, and Director of the Berkeley Evaluation and Assessment Research (BEAR) Center. His work spans a range of issues in measurement and assessment from the development of new statistical models for analysing measurement data, to the development of new assessments in diverse subject matter areas such as science education, patient-reported outcomes and child development.

References

Adams, R., and S. Khoo. 1996. Quest: The Interactive Test Analysis System. Version 2.1. Camberwell, Vic: ACER Press.
Adams, R., M. Wilson, and W. Wang. 1997. “The Multidimensional Random Coefficients Multinomial Logit Model.” Applied Psychological Measurement 21 (1): 1–23. doi:10.1177/0146621697211001.
Adams, R., M. Wu, and M. Wilson. 2012. ACER ConQuest 3.0. [Computer Programme]. Melbourne: ACER.
Allport, G. 1954. The Nature of Prejudice. Cambridge, MA: Addison-Wesley.
Andrich, D. A. 1978. “A Rating Formula for Ordered Response Categories.” Psychometrika 43: 561–573. doi:10.1007/BF02293814.
Ayala, R. 2013. “The IRT Tradition and Its Applications.” In The Oxford Handbook of Quantitative Methods, Oxford Library of Psychology edited by T. D. Little, 144–169. Oxford: Oxford University Press.
Benner, D., R. Schieder, H. Schluß, and J. Willems, eds. 2011. Religiöse Kompetenz als Teil öffentlicher Bildung: Versuch einer empirisch, bildungstheoretisch und religionspädagogisch ausgewiesenen Konstruktion religiöser Dimensionen und Anspruchsniveaus. Paderborn: Schöningh.
Bertram-Troost, G. 2011. “Investigating the Impact of Religious Diversity in Schools for Secondary Education: A Challenging but Necessary Exercise.” British Journal of Religious Education 33 (2): 271–283. doi:10.1080/01416200.2011.543614.
Boeck, P., and M. Wilson, eds. 2004. Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach. Statistics for Social Science and Public Policy. New York, NY:Springer. doi:10.1007/978-1-4757-3990-9.
Brockett, A., A. Village, and L. Francis. 2009. “Internal Consistency Reliability and Construct Validity of the Attitude toward Muslim Proximity Index (AMPI): A Measure of Social Distance.” British Journal of Religious Education 31 (3): 241–249. doi:10.1080/01416200903112417.
Cronbach, L. J. 1951. “Coefficient Alpha and the Internal Structure of Tests.” Psychometrika 16: 297–334. doi:10.1007/BF02310555.
Eagly, A., and S. Chaiken. 1993. The Psychology of Attitudes. Fort Worth, TX: Harcourt Brace Jovanovich College.
Embretson, S. 1996. “The New Rules of Measurement.” Psychological Assessment 8 (4): 341–349. doi:10.1037/1040-3590.8.4.341.
Embretson, S., and R. Reise. 2000. Item Response Theory for Psychologists. Multivariate Applications Book Series. Mahwah, NJ: Erlbaum.
European Commission. 2018. Special Eurobarometer 469 “Integration of Immigrants in the European Union”. Brussels.
Francis, L., G. Penny, and U. McKenna. 2016. “Does RE Work and Contribute to the Common Good in England?” In Young People’s Attitudes to Religious Diversity, edited by E. Arweck, 153–169. London, New York: Routledge Taylor & Francis.
Guttman, L. 1944. “A Basis for Scaling Qualitative Data.” American Sociological Review 9 (2): 139–150. doi:10.2307/2086306.
Guttman, L. 1950. “The Basis of Scalogram Analysis.” In Studies in Social Psychology in World War II: Vol. 4 Measurement and Prediction, edited by Samuel A. Stouffer, et al., 60–90. Princeton: Princeton University Press.
Hafez, K., and S. Schmidt. 2015. Die Wahrnehmung des Islams in Deutschland: Religionsmonitor – verstehen was verbindet. Gütersloh: Verlag Bertelsmann Stiftung.
Halm, D., and M. Sauer. 2017. Muslime in Europa: Integriert aber nicht akzeptiert? Religionsmonitor – verstehen was verbindet. Gütersloh: Bertelsmann Stiftung.
Hermissen, S. 2017. “Schülerereinstellungen zu religiöser und weltanschaulicher Diversität: Entwicklung eines Erhebungsinstruments auf der Basis der Item-Response-Theorie.” Theo-Web 16: 75–92.
Igrave, J. 2016. “Uniting Two Communities or Creating a Third Community? Research in a Northern Irish Integrated School.” In Arweck 2016: 61–76.
Konnemann, C., R. Asshoff, and M. Hammann. 2016. “Insights into the Diversity of Attitudes Concerning Evolution and Creation: A Multidimensional Approach.” Science Education 100 (4): 673–705. doi:10.1002/sce.21226.
Masters, G. 1982. “A Rasch Model for Partial Credit Scoring.” Psychometrika 47: 149–174. doi:10.1007/BF02296272.
Mislevy, R. 1987. “Exploiting Auxiliary Information about Examinees in the Estimation of Item Parameters.” Applied Psychological Measurement 11: 81–91. doi:10.1177/014662168701100106.
### Appendix A

**Table A1.** Variances and (disattenuated) correlations among the four dimensions attitudes towards Muslims, Christians, Jews and Non-religious (with standard error in parenthesis).

| Attitude towards Muslims | Attitude towards Christians | Attitude towards Jews | Attitude towards Non-religious |
|--------------------------|-----------------------------|-----------------------|-------------------------------|
| Attitude towards Muslims | 1                           |                       |                               |
| Attitude towards Christians | 0.836 (0.031)                 | 1                     |                               |
| Attitude towards Jews   | 0.858 (0.028)                 | 0.979 (0.010)         | 1                             |
| Attitude towards Non-religious | 0.685 (0.044)     | 0.767 (0.036)         | 0.877 (0.024) | 1 |
| Variances               | 1.002 (0.110)                | 0.997 (0.108)         | 1.173 (0.126) | 1.114 (0.122) |

**Table A2.** Difference among five religion categories in four dimensions (‘Muslim’ as the reference category), controlling for age and gender.

|                | Attitude towards Muslims | Attitude towards Christians | Attitude towards Jews | Attitude towards Non-religious |
|----------------|--------------------------|-----------------------------|-----------------------|-------------------------------|
| Christians     | −0.13 (0.17)              | 0.72 (0.16)***              | 0.66 (0.18)***        | 0.51 (0.16)***                |
| Jews           | −0.69 (0.28)*             | −0.01 (0.25)                | 0.52 (0.29)           | 0.60 (0.27)*                  |
| Non-religious  | −0.61 (0.21)**            | −0.02 (0.20)                | 0.20 (0.22)           | 1.19 (0.21)***                |
| Other          | −0.59 (0.37)              | 0.16 (0.35)                 | 0.18 (0.39)           | 0.44 (0.36)                   |

Note. Attitudes are statistically different from the reference category “Muslims” at p < .05 (*), p < .01 (**) and p < .001 (***)