# The Generalised Antisemitism (GeAs) Scale: Validity and Factor Structure

Daniel Allington, David Hirsh, and Louise Katz

## Abstract

This article validates the Generalised Antisemitism (GeAs) scale, which provides a measure of antisemitism consistent with the International Holocaust Remembrance Alliance Working Definition of Antisemitism (generally known as the IHRA Definition). The GeAs scale is comprised of two six-item subscales, each containing a balance of reverse-coded items: the Judeophobic Antisemitism (JpAs) subscale, comprised of antisemitic statements about Jews as Jews, and the Antizionist Antisemitism (AzAs) subscale, comprised of antisemitic statements about Israel and its supporters. Pre-registered tests of convergent-discriminant validity were carried out using a quota sample ($n = 602$), which was also used to test the pre-registered hypothesis of positive correlation between subscales. The latter was supported and shown to be robust to outliers, as well as to hold both among male and female respondents and among younger and older respondents. Test-retest reliability was measured using re-invitees from the first sample ($n = 428$). Data from larger samples of UK-resident adults (a quota sample balanced for age and gender, $n = 809$, and a representative random sample from a recruited panel, $n = 1853$) were used in a confirmatory factor analysis and in tests of measurement invariance. The findings provide further evidence that the GeAs scale is reliable and valid. The finding that improved fit is achieved by bifactor models featuring two group factors and a general factor is consistent with the view that statements characteristic of "old" and "new" antisemitism express a single underlying trait.

**Keywords:** antisemitism; antizionism; factor analysis; Israel; Jews; measurement

## INTRODUCTION

In their literature review of psychological research on antisemitism in the United States, Kaufman and colleagues note the need for an up-to-date scale: of the fifteen studies they reviewed, only one used an antisemitism scale developed less than forty-five years ago, and that was a slightly modified version of an older scale.[1] However, recent decades have seen the rise of what is called the "new antisemitism" or "new Judeophobia," that is, an extreme anti-Israel prejudice that closely resembles classic antisemitism and is in practice associated with targeting of

Jews and Jewish community institutions both in Israel and in the Diaspora.[2] Accordingly, the International Holocaust Remembrance Alliance Working Definition of Antisemitism, henceforth referred to as the IHRA Definition, explicitly recognizes that antisemitism may be expressed through "targeting of the state of Israel, conceived as a Jewish collectivity."[3]

Not only is there no standard measure of antisemitism that takes account of this development, there is none that features a balance of reverse-coded items. This is important both because meaningless acquiescence may produce spurious correlations between unbalanced scales[4]

and because respondents may interpret an unbalanced questionnaire as evidence of the researcher's opinion, in turn contributing to bias.[5] Levinson and Sanford's fifty-two-item scale, tested on a sample of seventy-seven female undergraduate students of psychology, features no reverse-coded items.[6] A shortened thirty-two-item version was found to achieve excellent internal reliability with a sample of 105 psychology students, although the researchers who developed it recommend that the wording of half the items should be reversed.[7] The Anti-Defamation League's eleven-item antisemitism index, originally developed by Glock and Stark,[8] is a more practical length for large-scale research,[9] but again features no reverse-coded items. Lastly, Selznick and Steinberg's[10] eighteen-item scale features only a small minority of reverse-coded items.[11] It may be observed that one of the above scales predates the establishment of the State of Israel, and that none of the others contains items relating to it, with the exception of "Jews are more loyal to Israel than to America", in the Selznick and Steinberg inventory. A number of studies have employed novel measures of more or less extreme anti-Israeli attitudes in order to test whether these predict anti-Jewish attitudes.[12]

However, while it would in principle be possible to combine these with measures of "old" antisemitism to produce an updated scale, none of these measures of anti-Israel prejudice employs a balance of positively and negatively keyed items either. A further issue is that some of the aforementioned scales have been tested only with small samples, while none has been subjected to factor analysis, and that little is known about their psychometric properties beyond measures of internal consistency.

The Generalised Antisemitism or GeAs scale was developed as a short, up-to-date inventory comprised of two equal-sized subscales, each with a balance of positively and negatively keyed items.[13] The Judeophobic Antisemitism or JpAs subscale was adapted from the inventory used in the annual Antisemitism Barometer survey conducted for Campaign against Antisemitism,[14] and the Antizionist Antisemitism or AzAs subscale was adapted from the identically named scale developed by Allington and Hirsh.[15] See table 1 for all items. The current article aims to validate this scale whilst also analysing its factor structure in order to answer the question of whether responses to questionnaire items indicative of "new" and "old" antisemitism express a

### TABLE 1. GeAs Scale Items

| Item | Reverse? | Text |
|------|----------|------|
| JpAs 1 | Y | Jewish people can be trusted just as much as other [nationality] people in business |
| JpAs 2 | Y | Jewish people are just as loyal to [nation] as other [nationality] people |
| JpAs 3 | Y | I am just as open to having Jewish friends as I am to having friends from other sections of [nationality] society |
| JpAs 4 | N | Compared to other groups, Jewish people have too much power in the media |
| JpAs 5 | N | Jewish people talk about the Holocaust just to further their political agenda |
| JpAs 6 | N | Jewish people chase money more than other people do |
| AzAs 1 | Y | I am comfortable spending time with people who openly support Israel |
| AzAs 2 | Y | Israel has a right to exist as a homeland for the Jewish people |
| AzAs 3 | Y | Israel is right to defend itself against those who want to destroy it |
| AzAs 4 | N | Israel and its supporters are a bad influence on our democracy |
| AzAs 5 | N | Israel can get away with anything because its supporters control the media |
| AzAs 6 | N | Israel treats the Palestinians like the Nazis treated the Jews |

single trait, a pair of related traits, or two unrelated traits combined with a single underlying trait (in which case, a bifactor model will achieve the best fit). Accordingly, four studies are presented below.

Four samples were used across those studies. These are referred to as the Test sample ($n = 602$, collected April 24–25, 2021) and the Retest sample ($n = 428$, collected May 7–10, 2021), both of which contained approximately equal numbers of male and female respondents based in the United Kingdom, the Balanced sample ($n = 809$, collected October 30–31 2020), also collected in the United Kingdom, which comprised approximately equal numbers of male and female respondents as well as approximately equal numbers of respondents under and over the age of twenty-five, and the Representative sample ($n = 1853$, collected December 16–17, 2020), which was commissioned as a nationally representative sample of the British population from the opinion research company, YouGov. All data were collected online, via YouGov's platform for the Representative sample and via the Qualtrics platform for the other three samples.

Study 1, which was pre-registered, tests the convergent-discriminant validity of the JpAs and AzAs subscales separately, using the Test sample. Study 2, which was also pre-registered, is a replication study testing for positive correlation between the subscales, again using the Test sample. For comparison, correlations between the two subscales in the Retest, Balanced, and Representative samples are also measured. Correlations within demographic subsamples of all four samples are also presented for information and as a form of exploratory analysis, and an unregistered sensitivity analysis is undertaken in order to measure the influence of outliers on the findings. It is thus the most analytically thorough test to date of the hypothesis that anti-Jewish and anti-Israeli attitudes are positively related, which follows from the theory of the "new antisemitism" (see above). Study 3 measures test-retest reliability of the complete scale and both of its subscales using the Retest sample. Lastly, Study 4

employs confirmatory factor analysis of data collected from the Balanced and Representative samples, and tests for measurement invariance across age and gender in both samples. It serves both to further establish the validity of the GeAs scale and to provide further empirical evidence about the relationship between anti-Jewish and anti-Israeli attitudes, beyond simple correlation.

## ETHICAL STATEMENT

Ethical and data protection guidelines at King's College, London, the University of Derby, and YouGov were followed. Ethical clearance was sought and received at King's College, London, MRA-19/20-20918. Informed consent was obtained from all participants.

## POWER ANALYSIS

Given a cut-off of $p = .010$ and a random sample for which $n = 600$, a true effect size of $r = .13$ provides eighty-percent power in a one-tailed test. It was thus anticipated that Study 1 and Study 2 would be adequately powered, provided that real effect sizes are not at the lower end of the conventional bracket for a weak effect.[16] Given the typically high levels of correlation for test-retest studies, power for Study 3 was anticipated to be virtually 100%. There is no standard approach to power analysis for the forms of analysis employed in Study 4.

## SAMPLE DEMOGRAPHICS

Equal quotas of male and female respondents were recruited for the Test and Retest samples. Equal quotas of male and female respondents aged eighteen to twenty-five and twenty-six or older were recruited for the Balanced sample. Information on gender was re-collected on the first screen of the questionnaire after consent was obtained, hence a small minority of participants categorised for analytic purposes as male were recruited as part of the female quota and vice versa. Nationally representative quotas for age, gender, and other demographic variables were used for the Representative sample, following YouGov standard procedures, and these variables were not re-measured through the primary data collection instrument.

For analytic purposes, members of all samples were divided into subsamples by both gender and age, using two age categories: eighteen to twenty-five and twenty-six or older. The decision to treat the age of twenty-five as the cut-off for the younger group was made with reference to the theoretical framework of emerging adulthood, which suggests that individuals within the eighteen to twenty-five age range can neither be described as adolescents nor as adults.[17] Instead, these emerging adults are entwined within a temporal period, which is epitomized by biopsychosocial changes that may incorporate unstable peer group relationships, cognitive elasticity, a precarious sense of self-identity, and behaviors of impulsivity.[18] Although the age of twenty-five as a cut-off point may appear capricious, Arnett[19] argues that established societal roles and a stable self-identity are attained from the age of twenty-six and over. It has furthermore been argued that from this age onward, the processing of socio-emotional information and reactions to emotional stimuli are more controlled, with decreases in negative emotionality and aggressiveness,[20] and that the emerging adult may become entwined within peer groups thereby avoiding exclusion and as such, the group's attitudes and behaviors become a salient source of information. Thus, although group norms are adhered to,[21] these norms may also play a role in the expression of prejudice and the enactment of prejudiced behavior.[22]

Demographic descriptive statistics for all samples and subsamples are presented in table 2, with mean and standard deviation for age and percentages for categorical demographic variables. The minority of respondents (seven members of the Test sample, nineteen members of the Retest sample, seven members of the Balanced sample, and no members of the Representative sample) who identified neither as male nor as female or who declined to answer the question on gender are counted neither in the Male nor in the Female column.

While the sample commissioned from YouGov was designed for representativeness and employed random sampling, it suffered from underrepresentation of male respondents (especially within the eighteen to twenty-five age group) due to non-response bias, and of members of other than white ethnic groups due to longstanding issues in the British polling industry.[23] However, in other respects, it was more representative than the three samples collected via Prolific. Members of the latter were noticeably young, with the mean age even of the over-twenty-six subsample of the Balanced sample being under forty-three, and very highly educated, with at least fifty percent of each sample being educated to college level.[24] Moreover, with regard to self-declared political orientation, there was a strong leftward skew in all three samples collected via Prolific, while something closer to balance was achieved in the Representative sample (5.1).

## SCALE VALIDATION

### Study 1: Convergent-Discriminant Testing of Subscales

#### Study 1: Introduction

Convergent-discriminant validity is of particular importance for the GeAs scale because the

**TABLE 2.** Demographic Descriptive Statistics

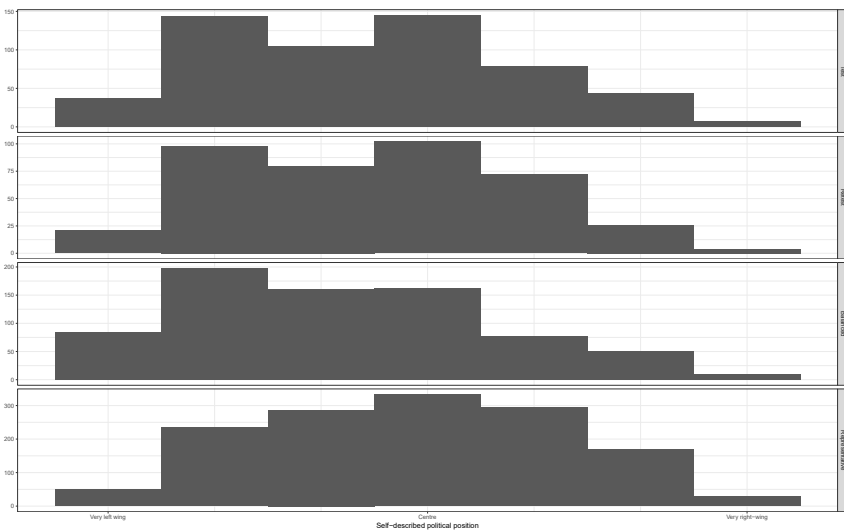| Sample, subsample | n | Age | | Gender (%) | | Ethnic group (%) | | Degree? (%) | |
|---|---|---|---|---|---|---|---|---|---|
| | | M | SD | Female | Male | White | Other | Yes | No |
| Test | 602 | 39.97 | 15.25 | 50 | 49 | 87 | 13 | 50 | 50 |
| Test, 18–25 | 136 | 21.12 | 2.25 | 46 | 51 | 71 | 29 | 39 | 61 |
| Test, 26+ | 466 | 45.47 | 12.85 | 51 | 49 | 92 | 8 | 53 | 47 |
| Test, Female | 300 | 40.47 | 14.79 | 100 | 0 | 91 | 9 | 50 | 50 |
| Test, Male | 295 | 39.79 | 15.69 | 0 | 100 | 84 | 16 | 49 | 51 |
| Retest | 428 | 42.45 | 14.88 | 51 | 48 | 88 | 12 | 51 | 49 |
| Retest, 18–25 | 67 | 21.24 | 2.24 | 48 | 48 | 71 | 28 | 39 | 61 |
| Retest, 26+ | 346 | 46.56 | 12.61 | 52 | 48 | 92 | 8 | 53 | 47 |
| Retest, Female | 211 | 42.36 | 14.14 | 100 | 0 | 90 | 9 | 52 | 48 |
| Retest, Male | 198 | 42.90 | 15.57 | 0 | 100 | 86 | 14 | 50 | 50 |
| Balanced | 809 | 32.42 | 13.72 | 50 | 49 | 82 | 17 | 53 | 47 |
| Balanced, 18–25 | 394 | 21.54 | 2.22 | 52 | 48 | 71 | 28 | 42 | 58 |
| Balanced, 26+ | 412 | 42.82 | 11.93 | 50 | 50 | 93 | 7 | 64 | 36 |
| Balanced, Female | 407 | 32.04 | 13.33 | 100 | 0 | 83 | 16 | 58 | 42 |
| Balanced, Male | 395 | 32.92 | 14.18 | 0 | 100 | 81 | 18 | 48 | 52 |
| Representative | 1853 | 51.85 | 16.68 | 59 | 41 | 94 | 5 | 30 | 70 |
| Representative, 18–25 | 144 | 21.98 | 2.11 | 76 | 24 | 78 | 22 | 31 | 69 |
| Representative, 26+ | 1709 | 54.37 | 14.82 | 58 | 42 | 95 | 4 | 30 | 70 |
| Representative, Female | 1095 | 49.60 | 17.06 | 100 | 0 | 94 | 6 | 31 | 69 |
| Representative, Male | 758 | 55.11 | 15.55 | 0 | 100 | 94 | 4 | 28 | 72 |



**FIGURE 1.** Political orientation of respondents.

question of the relationship between the constructs measured by its two subscales, that is, antipathy towards Jews and antipathy towards the Jewish state, is itself an area of political controversy. For this reason, fairly direct measures of attitudes to both were adopted, together with parallel measures of attitudes to other religious groups and to other countries. This part of the analysis was pre-registered in order to further eliminate doubt. It is, moreover, presented first among the analyses collected in this article (even though it was not carried out first) because it establishes the minimum conditions for the validity of the remaining analyses.

An innovative approach to convergent-discriminant testing was taken because of the lack of any pre-existing and generally accepted scales for the measurement of antizionist antisemitism or Judeophobic antisemitism. Had such scales existed, it would have made sense simply to test for correlation with them, with a strong correlation then being taken as an indication that the two subscales do indeed measure the same constructs. However, as things stand, all we would be able to measure is that the subscales correlate with other scales that have been proposed as measures of the same thing. Our approach was therefore instead to take single-item measures that correspond very intuitively to the constructs the subscales are designed to measure (that is, attitudes towards Jews and towards Israel), to set these alongside single-item measures that correspond just as intuitively to different but related constructs (that is, attitudes towards other religious minorities and towards other countries), and to compare the correlations. Because single-item measures inevitably measure attitudes in a rather crude way, high levels of correlation cannot be expected. However, clear differences in levels of correlation can be expected between items intuitively corresponding to the constructs that the subscales are intended to measure and items intuitively corresponding to different constructs. Thus, the pre-registered criteria for convergent-discriminant validity were defined not in

terms of an absolute threshold for the coefficients themselves, but in terms of (a) the statistical significance of the coefficients of correlation between subscales and single-item measures intuitively corresponding to the same intended construct and (b) the statistical significance of the difference between those coefficients of correlation and the coefficients of correlation between the same subscales and single-item measures intuitively corresponding to single-item measures intuitively corresponding to different but related constructs. The implicit theory here is that a scale measuring negative attitudes to a group in a relatively nuanced way should correlate negatively with a single item measuring positivity towards that group in a relatively un-nuanced way, and indeed should correlate more negatively with that item than with single items measuring positivity towards other groups in an equally crude way. While a scale measuring negative attitudes towards a nation state in a relatively nuanced way should correlate positively with a single item measuring positivity towards that country in a relatively un-nuanced way, and indeed should correlate more positively with that item than with single items measuring positivity towards other countries in a relatively un-nuanced way, even if the (inevitable) lack of nuance in single-item measures places limits on the absolute levels of correlation, which can reasonably be expected.

Slightly different approaches were taken to the two subscales in this regard, with a feeling thermometer question being used for the JpAs subscale and an ordinal scale question being used for the AzAs subscale. Feeling thermometers are a well-established measure of attitudes to religious groups. The regular American National Election study has used feeling thermometers to measure affective warmth towards Jews and Catholics since 1964, attitudes towards Christian Fundamentalists since 1988, attitudes towards Muslims since 2004, and attitudes towards Christians since 2008.[25] Psychological studies have used this question type to assess prejudice against diverse groups in diverse contexts, including

sexual minorities,[26] racial minorities,[27] and religious minorities.[28]

Although there would have been precedent for using feeling thermometers to measure attitudes to countries,[29] a different approach was taken for convergent-discriminant testing of the AzAs scale. As Harrison and Klaff argue, it is insufficient to define antisemitism solely as an "emotional disposition," as the word also properly denotes "a body of pseudo-explanatory political theory."[30] While the approach taken to convergent-discriminant testing of the JpAs subscale emphasised the emotional aspect of antisemitism in the form of affective warmth towards Jews, the approach take to convergent-discriminant testing of the AzAs subscale instead emphasised the pseudo-explanatory geopolitical aspect of antisemitism by measuring respondents' perceptions of various countries, including Israel, as threats to or defenders of world peace.

### Study 1: Hypotheses

The following hypotheses were pre-registered:[31]

H1.1 JpAs score correlates negatively with affective warmth towards Jewish people.

H1.2 Correlation between JpAs score and affective warmth towards Jewish people is more negative than correlation between JpAs score and affective warmth towards each of the following groups:

  H1.2.1 Christian people;
  H1.2.2 Muslim people;
  H1.2.3 Hindu people;
  H1.2.4 Sikh people;
  H1.2.5 Buddhist people;
  H1.2.6 People who are not religious.

H1.3 AzAs score correlates positively with assessment of Israel as a threat to world peace.

H1.4. Correlation between AzAs score and assessment of Israel as a threat to world peace is more positive than correlation between AzAs score and assessment of each of the following countries as a threat to world peace:

  H1.4.1 the United States;
  H1.4.2 the United Kingdom;

  H1.4.3 Russia;
  H1.4.4 China;
  H1.4.5 North Korea;
  H1.4.6 Iran.

The wording used for the feeling thermometer questions in the current study was closely adapted from that used in the 2008 British Social Attitudes study, which employed such questions to measure attitudes towards Catholics, Protestants, Muslims, and Buddhists, as well as towards the irreligious and the very religious.[32]

### Study 1: Analytic Methodology

The methodology to be used for testing each hypothesis was pre-registered:[33] H1.1 would be tested through calculation of the product-moment coefficient of correlation (given that feeling thermometer data is numeric), H1.3 through calculation of the rank-order coefficient of correlation (given that an ordinal measure of threat perception was used), and H1.2 and 1.4 would be tested through the Williams test of equality between dependent correlations (with product-moment and rank-order coefficients of correlation used as appropriate). In order to compensate for different response styles in the feeling thermometer questions (for example, where some respondents may use more of the 0–100 range than others), feeling thermometer responses for each respondent were standardised (that is, the respondent's thermometer rating for each religious group will be reduced by the same respondent's mean thermometer rating for all religious groups, and then divided by the standard deviation of his or her thermometer ratings for all religious groups), with this transformation also being pre-registered. (Note that where respondents gave the same thermometer rating to all groups, this calculation involved division by zero, and therefore produced missing data.)

A cut-off criterion of $p < .01$ was pre-registered for H1.1 and H1.3. For H1.2 and H1.4, the pre-registered approach was to use a cut-off criterion of $p < .05$ in combination with the Holm-Bonferroni method of adjustment for multiple comparisons.

*Study 1: Findings*

JpAs was found to be negatively correlated with subjective warmth towards Jewish people, $r(484) = -.27$, $p <.001$, 95% CI [–1.00, –.20] (one-tailed). H1.1 was thus supported, that is, the JpAs subscale was found to have convergent validity. By contrast, JpAs correlated positively with subjective warmth towards Christians, and was uncorrelated or very weakly correlated with subjective warmth towards other religious groups: see table 3 for coefficients of correlation with two-tailed ninety-five-percent confidence intervals.

For purposes of discriminant reliability testing, one-tailed Williams tests were used to test for differences between correlation coefficients between JpAs and subjective warmth towards Jewish people and correlations coefficients between JpAs and subjective warmth towards other religious groups. For every other group, the correlation with JpAs was less negative than it was for Jewish people, $p = .002$ or $p <.001$ (see table 4). This represents such a high level of statistical significance that the pre-registered use of a more lenient level of alpha combined with the Holm-Bonferroni method was unnecessary: H1.2.1–6 are supported, that is, the JpAs subscale was found to have discriminant validity.

AzAs was found to be positively correlated with perception of Israel as a threat to world peace, $r(600) = .48$, $p <.001$, 95% CI [.43, 1.00] (one-tailed). H1.3 was thus supported, that is, the AzAs scale was found to have convergent validity. AzAs was also positively correlated—although less strongly—with perception of Israel's allies, the United States and the United Kingdom, as threats to world peace, and was uncorrelated with perception of the other countries as threats to world peace, except for North Korea and Iran, where there was a very weak negative correlation. See table 5 for coefficients of correlation with two-tailed ninety-five-percent confidence intervals.

For purposes of discriminant reliability testing, one-tailed Williams tests were used to

**TABLE 3.** Product-moment Correlations between Jpas and Subjective Warmth towards Non-Jewish Religious Groups

| Group | Est. | Low | High | p |
|---|---|---|---|---|
| Christian | .33 | .24 | .40 | <.001 |
| Muslim | -.02 | -.11 | .07 | .636 |
| Sikh | -.09 | -.18 | .00 | .047 |
| Hindu | -.04 | -.13 | .05 | .420 |
| Buddhist | -.05 | -.14 | .04 | .247 |
| Not religious | -.02 | -.11 | .07 | .617 |

**TABLE 4.** Williams Tests of the Hypothesis that Correlation between JpAs and Subjective Warmth towards Non-Jewish Religious Groups is Less Negative than it is for Subjective Warmth Towards Jewish People

| Group | t | p |
|---|---|---|
| Christian | -9.68 | <.001 |
| Muslim | -3.67 | <.001 |
| Sikh | -2.85 | .002 |
| Hindu | -3.66 | <.001 |
| Buddhist | -3.33 | <.001 |
| Not religious | -3.65 | <.001 |

**TABLE 5.** Rank-order Correlations between AzAs and Perception of Countries other than Israel as a Threat to World Peace

| Country | Est. | Low | High | p |
|---|---|---|---|---|
| US | .32 | .24 | .39 | <.001 |
| UK | .29 | .21 | .36 | <.001 |
| Russia | .02 | -.07 | .09 | .688 |
| China | .03 | -.05 | .12 | .443 |
| North Korea | -.09 | -.18 | -.02 | .022 |
| Iran | -.10 | -.18 | -.02 | .011 |

test for differences between correlation coefficients between AzAs and perception of Israel as a threat to world peace and correlations

coefficients between JpAs and perception of other countries as a threat to world peace. For every other country, the correlation with AzAs was less positive than it is for Israel, $p$ <.001 (see table 6). Again, this represents such a high level of statistical significance that the pre-registered use of a more lenient level of alpha combined with the Holm-Bonferroni method was unnecessary: H1.4.1–6 were supported, that is, the AzAs subscale was found to have discriminant validity.

### Study 1: Discussion

JpAs correlated more negatively with subjective warmth towards Jewish people than with subjective warmth towards any other religious group. This was a strong indication that it is indeed a measure of anti-Jewish attitudes. The positive correlation with subjective warmth towards Christians was unexpected, and no explanation for it is attempted here. AzAs correlated more positively with perception of Israel as a threat to world peace than with perception of any other country as a threat to world peace. The positive correlation with perception of the United States and the United Kingdom as a threat to world peace suggests that anti-Israeli attitudes may form part of a wider politics—Wistrich, for example, writes of "a loose and shifting coalition of red-brown-green bigotry focused against both America and Israel"[34]—but the finding that the

correlation with perception of Israel as a threat was so much stronger is evidence that AzAs is indeed a measure of specifically anti-Israeli attitudes. Both subscales of the GeAs scale can thus be assumed to be working as intended. This provides support for the view that relationships between the subscales—investigation of which is the focus of the following study—may reasonably be taken to reflect relationships between the constructs that they are intended to measure.

### Study 2: Correlation between Subscales

Study 2 attempts to replicate the repeated empirical finding that heightened anti-Jewish attitudes are predicted by heightened anti-Israeli attitudes.[35] Accordingly, a one-tailed correlation test was pre-registered with regard to the Test sample. However, in the interests of maximum informativeness, correlations (with two-tailed confidence intervals and $p$-values) are also calculated for the JpAs and AzAs subscales in the remaining three samples, and in demographic subsamples of all samples.

### Study 2: Hypothesis

Given the repeated finding of a positive relationship between anti-Jewish and anti-Israeli sentiment or opinion, the following hypothesis was pre-registered with regard to the Test sample:[36]

> H2.1 JpAs and AzAs are positively correlated.

### Study 2: Analytic Methodology

The analytic methodology of testing H2.1 through calculation of the product-moment coefficient of correlation was pre-registered.[37] For informational purposes, product-moment correlations for the two subscales were also calculated for the remaining three samples, and for demographic subsamples of all four samples by age and gender. (Here and elsewhere in this article, scores for the Retest sample refer to the retest scores themselves, except where otherwise indicated.)

TABLE 6. Williams Tests of the Hypothesis that Correlation between AzAs and Perception of Countries other than Israel as a Threat is Less Positive than it is for Perception of Israel as a Threat

| Country | $t$ | $p$ |
| --- | --- | --- |
| US | 3.93 | <.001 |
| UK | 4.38 | <.001 |
| Russia | 9.89 | <.001 |
| China | 9.44 | <.001 |
| North Korea | 11.61 | <.001 |
| Iran | 12.56 | <.001 |

## Study 2: Findings

Mean GeAs, JpAs, and AzAs scores were lowest in the Balanced sample, not only for the sample as a whole, but also for all subsamples. (It will be seen that findings for that sample were slightly anomalous in certain other respects.) In every sample, JpAs scores were higher among male respondents than among females while AzAs scores were higher among female respondents than among males: differences that tended to even out with regard to GeAs scores. Descriptive statistics for GeAs and its two subscales are presented in table 7, for all four samples and for age and gender-based subsamples thereof. Figure 2 presents histograms for the same variables across the same samples.

TABLE 7. GeAs, JpAs, and AzAs, Across Samples and Subsamples: Mean, Standard Deviation, Skewness, and Kurtosis

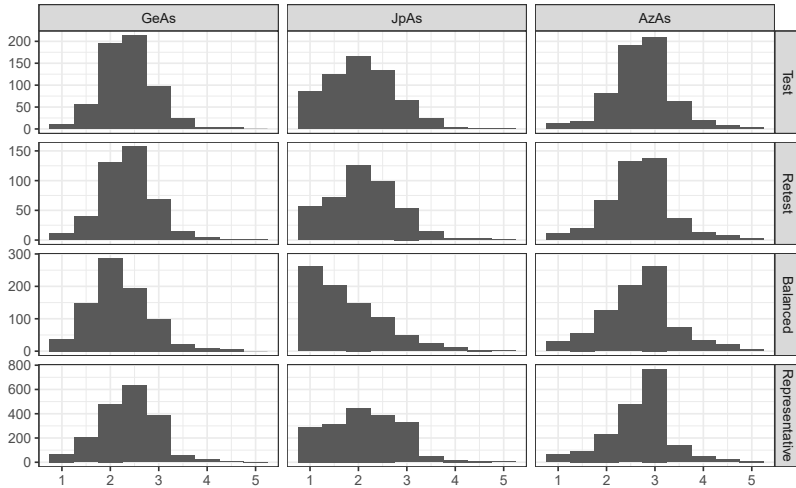| Sample, subsample | | GeAs | | | | JpAs | | | | AzAs | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | M | SD | *SK* | *κ* | M | SD | *SK* | *κ* | M | SD | *SK* | *κ* |
| Test | 602 | 2.41 | 0.54 | 0.54 | 4.55 | 2.07 | 0.69 | 0.51 | 3.35 | 2.75 | 0.62 | 0.15 | 4.35 |
| Test, 18–25 | 136 | 2.38 | 0.51 | 0.98 | 4.57 | 1.87 | 0.67 | 0.69 | 2.98 | 2.88 | 0.61 | 0.64 | 4.15 |
| Test, 26+ | 466 | 2.41 | 0.55 | 0.43 | 4.55 | 2.12 | 0.68 | 0.47 | 3.53 | 2.70 | 0.62 | 0.02 | 4.30 |
| Test, Female | 300 | 2.38 | 0.45 | 0.37 | 4.00 | 1.98 | 0.61 | 0.28 | 2.60 | 2.78 | 0.54 | 0.25 | 4.81 |
| Test, Male | 295 | 2.44 | 0.62 | 0.50 | 4.15 | 2.16 | 0.74 | 0.51 | 3.38 | 2.71 | 0.70 | 0.15 | 3.85 |
| Retest | 428 | 2.40 | 0.55 | 0.36 | 4.46 | 2.12 | 0.69 | 0.41 | 3.31 | 2.68 | 0.65 | 0.20 | 4.13 |
| Retest, 18–25 | 67 | 2.45 | 0.46 | 0.65 | 3.77 | 1.97 | 0.66 | 0.45 | 2.76 | 2.93 | 0.58 | 0.48 | 3.38 |
| Retest, 26+ | 346 | 2.40 | 0.56 | 0.49 | 4.56 | 2.15 | 0.69 | 0.44 | 3.53 | 2.65 | 0.63 | 0.36 | 4.29 |
| Retest, Female | 211 | 2.41 | 0.48 | 0.68 | 6.10 | 2.06 | 0.63 | 0.38 | 4.05 | 2.77 | 0.55 | 0.75 | 5.20 |
| Retest, Male | 198 | 2.41 | 0.60 | 0.37 | 3.59 | 2.19 | 0.74 | 0.40 | 2.89 | 2.63 | 0.70 | 0.21 | 3.34 |
| Balanced | 809 | 2.22 | 0.62 | 0.71 | 4.06 | 1.77 | 0.77 | 1.13 | 4.09 | 2.68 | 0.74 | 0.14 | 3.47 |
| Balanced, 18–25 | 394 | 2.25 | 0.59 | 0.65 | 4.01 | 1.74 | 0.73 | 0.98 | 3.35 | 2.76 | 0.75 | 0.38 | 3.65 |
| Balanced, 26+ | 412 | 2.20 | 0.65 | 0.77 | 4.07 | 1.80 | 0.81 | 1.21 | 4.43 | 2.60 | 0.73 | -0.13 | 3.04 |
| Balanced, Female | 407 | 2.17 | 0.54 | 0.62 | 4.26 | 1.61 | 0.65 | 1.13 | 3.81 | 2.73 | 0.71 | 0.19 | 3.85 |
| Balanced, Male | 395 | 2.28 | 0.69 | 0.63 | 3.55 | 1.95 | 0.85 | 0.95 | 3.61 | 2.62 | 0.77 | 0.11 | 3.11 |
| Representative | 1853 | 2.40 | 0.59 | 0.17 | 3.82 | 2.12 | 0.75 | 0.40 | 3.09 | 2.69 | 0.65 | -0.20 | 4.21 |
| Representative, 18–25 | 144 | 2.54 | 0.58 | 0.16 | 2.80 | 2.18 | 0.84 | 0.04 | 2.10 | 2.91 | 0.61 | 0.60 | 5.56 |
| Representative, 26+ | 1709 | 2.39 | 0.59 | 0.17 | 3.91 | 2.11 | 0.74 | 0.43 | 3.22 | 2.67 | 0.65 | -0.25 | 4.07 |
| Representative, Female | 1095 | 2.41 | 0.53 | -0.06 | 3.43 | 2.05 | 0.70 | 0.26 | 2.55 | 2.76 | 0.57 | -0.23 | 5.16 |
| Representative, Male | 758 | 2.40 | 0.67 | 0.33 | 3.71 | 2.22 | 0.80 | 0.44 | 3.31 | 2.59 | 0.75 | 0.00 | 3.40 |

**Figure 2.** Scores for GeAs, JpAs, and AzAs.

AzAs and JpAs were found to be positively correlated, $r(600) = .35$, $p <.001$ 95% CI [.29, 1.00] (one-tailed). H2.1 was thus supported. See table 8 for coefficients of correlation with two-tailed 95% confidence intervals for all samples except for the Test sample, and for subsamples by age and gender within all four samples (inferential statistics for the Test sample as a whole were reported in this paragraph but excluded from the table because the test was pre-registered as one-tailed). See Figure 3 for scatterplots with smoothed conditional means and two-tailed ninety-five-percent confidence intervals for all four samples. Especially with the two larger samples, the conditional means fall close to a straight line, providing intuitive support for the idea of a linear relationship between attitudes to Jews and attitudes to Israel.

Coefficients of correlation were very consistent across the Test, Retest, and Balanced samples: $r = .35$, $r = .37$, and $r = .34$ respectively. The coefficient of correlation was higher for the Representative sample, that is, $r = .42$, with the lower bound of the ninety-five-percent confidence interval for the Representative sample being higher than the point estimates for all of the other samples. However, almost identical



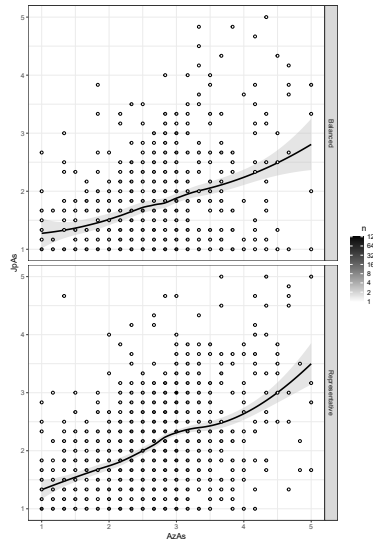**Figure 3.** JpAs vs AzAs, with locally estimated scatterplot smoothing and ninety-five-percent confidence intervals (two-tailed).

correlations were found among males, both among participants aged twenty-six and older, and among eighteen- to twenty-five-year-olds, regardless of whether one looks at the Test, Balanced, or Representative samples: it was only among female respondents that dramatically

TABLE 8. Product-moment Correlations, JpAs and AzAs

| Sample, subsample | DF | Est. | Low | High | p |
|---|---|---|---|---|---|
| Test* | 600 | .35 | | | |
| Test, 18–25 | 134 | .25 | .09 | .41 | .003 |
| Test, 26+ | 464 | .41 | .33 | .49 | <.001 |
| Test, Female | 298 | .22 | .11 | .33 | <.001 |
| Test, Male | 293 | .46 | .37 | .55 | <.001 |
| Retest | 423 | .37 | .29 | .45 | <.001 |
| Retest, 18–25 | 65 | .11 | -.13 | .34 | .376 |
| Retest, 26+ | 341 | .42 | .33 | .50 | <.001 |
| Retest, Female | 206 | .33 | .20 | .44 | <.001 |
| Retest, Male | 196 | .39 | .26 | .50 | <.001 |
| Balanced | 806 | .34 | .28 | .40 | <.001 |
| Balanced, 18–25 | 392 | .27 | .17 | .36 | <.001 |
| Balanced, 26+ | 410 | .42 | .34 | .50 | <.001 |
| Balanced, Female | 405 | .24 | .15 | .33 | <.001 |
| Balanced, Male | 393 | .47 | .39 | .54 | <.001 |
| Representative | 1851 | .42 | .38 | .46 | <.001 |
| Representative, 18–25 | 142 | .28 | .12 | .43 | .001 |
| Representative, 26+ | 1707 | .43 | .39 | .47 | <.001 |
| Representative, Female | 1093 | .38 | .33 | .43 | <.001 |
| Representative, Male | 756 | .50 | .44 | .55 | <.001 |

*Confidence interval and p-value of correlation reported separately because pre-registered as one-tailed.

different coefficients of correlation were observed between the Representative sample and the other samples. This suggests that there might be a problem of representativeness with regard to female participants reachable through the Prolific platform, at least with regard to antisemitism.

Comparing correlation coefficients across the demographic subsamples revealed that these were positive in every single case, and that—with the exception of the smallest subsample, that is, the eighteen-to-twenty-five-year-old subsample of the Retest sample, where the correlation was positive but statistically insignificant—all of these correlations were very highly statistically significant. In every sample, correlations were

higher among male respondents than among female respondents, although gender differences were much less pronounced in the Representative sample than in the Test or Balanced samples.

Sensitivity analysis was carried out in order to measure the possible effect of outliers on the above findings, through two-tailed comparison of correlation coefficients before and after their removal (where an outlier is defined as an observation for which the Cook's distance is greater than three times the mean). In the Test sample, there were thirty outliers, removal of which made no noticeable difference to the correlation. In the Retest sample, there were twenty-six outliers, removal of which increased the apparent correlation to $r = .44$. In the Balanced sample,

there were fifty-eight outliers, removal of which made no noticeable difference to the correlation. In the Representative sample, there were ninety-seven outliers, removal of which increased the apparent correlation to $r = .47$.[38] See table 9 for full breakdown with ninety-five-percent confidence intervals (two-tailed): the only statistically significant difference was the increase in the correlation coefficient for female members of the Representative sample (which had the effect of reducing the male-female gap).

### Study 2: Discussion

Given that Study 2 was in effect a replication of multiple previous studies carried out over the past fifteen years, it is worth comparing the findings with those earlier studies that also reported findings in the form of correlation coefficients.

Some previous studies reported correlations between numbers of antisemitic statements agreed with rather than correlations for Likert scores but this did not seem to affect the results greatly: for example, the product-moment correlation between numbers of JpAs and AzAs items agreed with (or disagreed with, for reverse-coded items) in the Representative sample was $r = .41$ (before removal of outliers), which was effectively identical to that reported above.

The strongest correlation between Likert scores for anti-Jewish and anti-Israeli sentiment, that is, $r = .61$, was measured in a small ($n = 194$) convenience sample by Baum and Nakazawa using scales with no reverse-coded items.[39] The possible role of satisficing behaviors in inflating correlations between unbalanced scales has been noted by researchers.[40] Using partially balanced

TABLE 9. Outlier Sensitivity of Product-moment Correlations, JpAs and AzAs

| Sample, subsample | Outliers* | Correlations following removal | | | | | Effect of removal | | |
| | | DF | Est. | Low | High | p | Δ | z | p |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Test | 30 | 570 | .36 | .28 | .43 | <.001 | .00 | 0.09 | .930 |
| Test, 18–25 | 9 | 125 | .18 | .00 | .34 | .046 | -.08 | 0.64 | .520 |
| Test, 26+ | 26 | 438 | .43 | .35 | .51 | <.001 | .02 | 0.38 | .701 |
| Test, Female | 19 | 279 | .26 | .15 | .37 | <.001 | .04 | 0.55 | .580 |
| Test, Male | 14 | 279 | .45 | .35 | .54 | <.001 | -.01 | 0.15 | .877 |
| Retest | 26 | 397 | .44 | .35 | .51 | <.001 | .06 | 1.10 | .271 |
| Retest, 18–25 | 7 | 58 | -.03 | -.28 | .22 | .810 | -.14 | 0.78 | .435 |
| Retest, 26+ | 19 | 322 | .43 | .34 | .51 | <.001 | .01 | 0.21 | .836 |
| Retest, Female | 8 | 198 | .37 | .24 | .48 | <.001 | .04 | 0.48 | .631 |
| Retest, Male | 16 | 180 | .45 | .32 | .56 | <.001 | .06 | 0.69 | .492 |
| Balanced | 58 | 748 | .34 | .27 | .40 | <.001 | .00 | 0.09 | .931 |
| Balanced, 18–25 | 26 | 366 | .26 | .17 | .36 | <.001 | .00 | 0.07 | .945 |
| Balanced, 26+ | 29 | 381 | .41 | .32 | .49 | <.001 | -.01 | 0.13 | .893 |
| Balanced, Female | 21 | 384 | .23 | .14 | .32 | <.001 | -.01 | 0.15 | .883 |
| Balanced, Male | 29 | 364 | .47 | .39 | .55 | <.001 | .00 | 0.06 | .956 |
| Representative | 97 | 1754 | .47 | .43 | .50 | <.001 | .05 | 1.85 | .064 |
| Representative, 18–25 | 9 | 133 | .38 | .22 | .51 | <.001 | .10 | 0.89 | .372 |
| Representative, 26+ | 86 | 1621 | .47 | .43 | .51 | <.001 | .04 | 1.39 | .164 |
| Representative, Female | 51 | 1042 | .45 | .40 | .50 | <.001 | .07 | 1.97 | .049 |
| Representative, Male | 42 | 714 | .53 | .48 | .58 | <.001 | .04 | 0.91 | .361 |

* Defined as observations for which the Cook's distance is greater than three times the mean.

scales and a representative sample, Staetsky reported a weaker, but still substantial correlation between numbers of anti-Jewish and anti-Israeli attitudes, $r = .48$. Both of those correlations were higher than those found in this study—although, following removal of outliers, the correlation coefficient for the Representative sample was effectively identical with Staetsky's. Beattie,[41] however, reported a correlation of $r = .36$ between numbers of anti-Jewish and extreme anti-Israeli sentiments (as opposed to a correlation of $r = .15$ for moderate anti-Israeli sentiments), which was slightly higher than that found for the Test sample but considerably lower than that found for the Representative sample. As correlations measured in the current study were substantially lower in all of the self-selecting samples recruited via the Prolific platform than they were in the Representative sample collected by YouGov, it is worth noting that Beattie also used a self-selecting sample recruited via the similar M-Turk platform. It is plausible that characteristic questionnaire-completion behaviors on the part of 'professional' questionnaire respondents reachable through platforms such as M-Turk and Prolific may account for these differences, whether through lower attention,[42] which can be assumed to increase statistical noise, or greater sophistication,[43] which might potentially lead to more successful disguising of socially undesirable attitudes.

As noted above, sensitivity analysis suggested that the findings reported here are robust to outliers. However, it raises the possibility that the apparently novel finding of lower correlation among female respondents may have been very slightly exaggerated by the effect of outliers in the female subsample, removal of which increased the correlation among members of this group by a small but significant amount.

## Study 3: Test-Retest Reliability

Members of the Retest sample responded to the GeAs items twice: the first time, with other members of the Test sample; and the second time, sixteen or seventeen days later. When Retest scores are referred to elsewhere in this article, this refers to the second data collection only. In the current study, scores on both occasions are compared to judge stability over time.

### Study 3: Analytic Methodology

Product-moment correlations were used as a two-tailed test of the relationship between the same respondents' scores in the Test and Retest samples.

### Study 3: Findings

Figure 4 shows changes in scores for the GeAs scale and for the JpAs and AzAs subscales. In ninety percent of cases, absolute difference
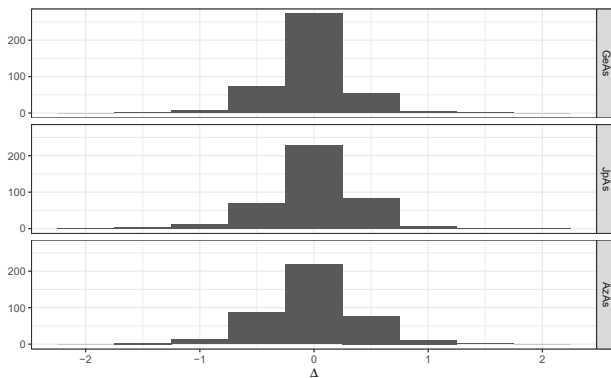


FIGURE 4. Differences in scores after sixteen or seventeen days.

between the first and second scores was no more than 0.50 of a point (or twelve percent of the total range) for GeAs, 0.67 of a point (or seventeen percent of the total range) for JpAs, and 0.67 of a point (or seventeen percent of the total range) for AzAs. Retested respondents' GeAs scores on the two occasions were very strongly correlated, $r(408) = .84$, $p <.001$ 95% CI [.81, .87] (two-tailed). There was also a very strong correlation between repeated scores both on the JpAs subscale, $r(408) = .82$, $p <.001$ 95% CI [.78, .85] (two-tailed) and on the AzAs subscale, $r(408) = .80$, $p <.001$ 95% CI [.76, .83] (two-tailed).

### Study 3: Discussion

Cicchetti suggests .80 as a threshold for a "good" level of clinical significance with regard to test-retest reliability.[44] This threshold is met by the GeAs scale as a whole, as well as by the JpAs and AzAs subscales considered individually, which suggests that responses to the GeAs items are stable over time.

## Study 4: Confirmatory Factor Analysis and Analysis of Invariances

In the final study, confirmatory factor analysis is used to investigate the relationship between the items comprising the JpAs and AzAs subscales of the GeAs scale, and to provide further measures of validity. This was done through comparison of different models: a one-factor model, which embodies the hypothesis that there is no distinction to be made between the JpAs and AzAs items (reflecting the assumption that "old" and "new" antisemitism are identical), a two-factor model embodying the hypothesis that the items making up the JpAs and AzAs subscales measures separate although potentially correlated traits (reflecting the assumption that "old" and "new" antisemitism are separate, but may exist in a relationship with one another), and a series of bifactor models. These embody variants of the hypothesis that the items of the JpAs and AzAs subscales measure distinct traits, but that all

GeAs items together additionally measure a general trait.

Use of bifactor models also permitted modelling of acquiescence bias as an additional general factor, orthogonal to all substantive factors, onto which the loadings of all items are fixed at 1 or −1, as appropriate.[45] In acknowledgement of the argument that a given individual's respondent's engagement in satisficing behaviors will vary according to topic interest and topic knowledge,[46] acquiescence bias was modelled first with a single factor for all items and then with separate factors for JpAs and for AzAs.

### Study 4: Hypotheses

The following five models constitute the hypotheses to be tested in Study 4 through confirmatory factor analysis. Acquiescence bias factor loadings are fixed to 1 or −1, and, by convention, the loading of the first indicator for each factor is fixed to 1; in all other cases, fixed loadings or correlations are fixed to 0. (Note that negatively keyed items were reverse-coded before analysis began.) In the two-factor model, the factors were allowed to correlate. In the bifactor models, all factors were constrained to be orthogonal to one another apart from the acquiescence bias factors, which were allowed to correlate with one another, although not with any other factors. Four of the five models are visualized in Figure 5 (the models for H5.4 and H5.5 are too similar to require separate diagrams), wherein curved lines represent correlations, straight lines represent factor loadings, solid lines represent correlations or loadings whose coefficients are to be estimated through fitting to the data, and dashed lines represent correlations or loadings whose coefficients are fixed.

> H5.1 A one-factor model for all items (figure 5, top left).
> H5.2 Separate factors for JpAs and GeAs items (figure 5, top right).
> H5.3 A bifactor model with group factors for JpAs and AzAs and a general factor (figure 5, bottom left).
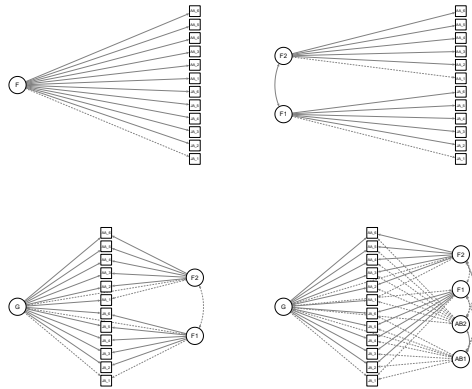
**FIGURE 5.** Factor models for the GeAs scale: one-factor (top left), two-factor (top right), bifactor (bottom left), bifactor with two acquiescence bias factors (bottom right).(bottom left), bifactor with two acquiescence bias factors (bottom right).

H5.4 A bifactor model (as in H5.3) with an acquiescence bias factor (not visualized).

H5.5 A bifactor model (as in H5.3) with two acquiescence bias factors (figure 5, bottom right).

### Study 4: Analytic Methodology

All models were fitted using diagonally weighted least squares as an estimator. Bifactor models are notoriously difficult to fit, because it frequently happens that solutions are found to fit equally well, resulting in non-convergence. When working with entire samples, the approach employed here was to fit the most complex model with a relaxed level of relative tolerance ($1 \times 10^{-5}$), extract the loadings, and then use those as starting values when fitting the same model at a more stringent level of relative tolerance ($1 \times 10^{-10}$). When working with subsamples (including for the analysis of invariances), even this approach did not always result in convergence, and therefore models were fitted in a single pass, using a moderate level of relative tolerance ($1 \times 10^{-7}$).

Root mean square error of approximation (RMSEA) is a measure of how close a particular model is to a hypothetical model that would

perfectly fit the data. It is conventionally accepted that RMSEA <.08 indicates an acceptable fit, while RMSEA <.06 or <.05 indicates a good fit.[47] However, Xia and Yang observe that these thresholds are based on intuition or—at best—on calculations involving assumptions that are not typically met with psychometric data, and argue that conventional cut-off points for measures of model fit should be discarded, with measures such as RMSEA instead being treated as "diagnostic tools for model improvement."[48] Thus, the five hypotheses for Study 4 were tested through comparison of RMSEA scores with ninety-five-percent confidence intervals (two-tailed) for fit to the Balanced and Representative samples. The comparative fit index (CFI), Tucker-Lewis index (TLI), and chi-square test are also provided for informational purposes.

Having established the best-fitting model, it is necessary to test for measurement invariance in order to determine whether items play comparable roles within the model across demographic groups. Testing for configural invariance involves fitting a model to various groups while allowing factor loadings and item intercepts to vary across groups. Testing for metric invariance involves the repeating the procedure while

constraining factor loadings to be equivalent, and testing for scalar invariance additionally involves constraining item intercepts to be equivalent. The conventional approach is to calculate the difference in fit indices between a model and the preceding less constrained model, interpreting a difference no larger than .010 for CFI and no larger than .015 for RMSEA as evidence that a particular level of invariance has been attained.[49] For the purposes of invariance testing, grouping by age is here accomplished by dividing respondents into the same two age groups used throughout this paper, that is, eighteen to twenty-five and twenty-six or over, while grouping by gender is accomplished by dividing respondents identifying as male or female into groups accordingly. Respondents who did not identify with one of the two canonical genders could not be included in the analysis of invariance by gender, as they were too few in number to constitute a subsample in their own right. However, they were included in the analysis of invariance by age.

### Study 4: Findings

Inter-item rank-order correlations for the entire GeAs scale are presented in table 10 (for the balanced sample) and table 11 (for the representative sample). Table 12 compares measures of fit for the five hypothetical models. In both samples, fit improved on all measures as we moved from a one-factor model to a two-factor model and then to a bifactor model, while modelling for acquiescence bias improved fit further still. The finding that a bifactor model fits the data better than a two factor model supports the view that there is a single, latent trait or factor, which is expressed in antisemitic attitudes both to Jews and to Israel.

Use of a single acquiescence bias factor brought about a notable improvement in fit in both samples, while the use of a second acquiescence bias factor improved fit slightly further in the Representative sample. The conventional RMSEA threshold for good fit of .05 or less was exceeded here by the bifactor model with two acquiescence bias factors when fitted to the Representative sample, but not by any model fitted to the Balanced sample, for which only the less stringent threshold of .08 was passed. However, as noted above, we are here using fit indices as a guide to model improvement rather than as an absolute measure of model quality. Fit indices for this model across demographic subsamples are presented in table 13 (note that the model was fitted to subsamples at a more relaxed

**TABLE 10.** Inter-item Rank-order Correlations: Balanced Sample

|  | Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | JpAs 1 |  | .58 | .52 | .63 | .46 | .41 | .26 | .41 | .18 | .09 | .21 | .09 |
| 2. | JpAs 2 | .58 |  | .54 | .56 | .45 | .45 | .25 | .32 | .14 | .10 | .15 | .02 |
| 3. | JpAs 3 | .52 | .54 |  | .53 | .61 | .65 | .21 | .29 | .10 | .12 | .23 | .10 |
| 4. | JpAs 4 | .63 | .56 | .53 |  | .48 | .40 | .14 | .28 | .06 | .05 | .11 | .00 |
| 5. | JpAs 5 | .46 | .45 | .61 | .48 |  | .51 | .23 | .24 | .10 | .13 | .20 | .06 |
| 6. | JpAs 6 | .41 | .45 | .65 | .40 | .51 |  | .16 | .23 | .11 | .16 | .22 | .13 |
| 7. | AzAs 1 | .26 | .25 | .21 | .14 | .23 | .16 |  | .50 | .46 | .41 | .37 | .36 |
| 8. | AzAs 2 | .41 | .32 | .29 | .28 | .24 | .23 | .50 |  | .43 | .36 | .32 | .29 |
| 9. | AzAs 3 | .18 | .14 | .10 | .06 | .10 | .11 | .46 | .43 |  | .36 | .21 | .33 |
| 10. | AzAs 4 | .09 | .10 | .12 | .05 | .13 | .16 | .41 | .36 | .36 |  | .43 | .51 |
| 11. | AzAs 5 | .21 | .15 | .23 | .11 | .20 | .22 | .37 | .32 | .21 | .43 |  | .55 |
| 12. | AzAs 6 | .09 | .02 | .10 | .00 | .06 | .13 | .36 | .29 | .33 | .51 | .55 |  |

TABLE 11. Inter-item Rank-order Correlations: Representative Sample

| | Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | JpAs 1 | | .72 | .70 | .57 | .58 | .52 | .23 | .29 | .20 | .28 | .29 | .21 |
| 2. | JpAs 2 | .72 | | .62 | .55 | .56 | .47 | .25 | .30 | .19 | .28 | .28 | .22 |
| 3. | JpAs 3 | .70 | .62 | | .47 | .57 | .45 | .22 | .27 | .17 | .23 | .22 | .13 |
| 4. | JpAs 4 | .57 | .55 | .47 | | .62 | .62 | .21 | .24 | .15 | .32 | .39 | .25 |
| 5. | JpAs 5 | .58 | .56 | .57 | .62 | | .57 | .21 | .28 | .16 | .30 | .32 | .21 |
| 6. | JpAs 6 | .52 | .47 | .45 | .62 | .57 | | .14 | .18 | .07 | .23 | .30 | .19 |
| 7. | AzAs 1 | .23 | .25 | .22 | .21 | .21 | .14 | | .46 | .46 | .48 | .37 | .42 |
| 8. | AzAs 2 | .29 | .30 | .27 | .24 | .28 | .18 | .46 | | .60 | .41 | .35 | .27 |
| 9. | AzAs 3 | .20 | .19 | .17 | .15 | .16 | .07 | .46 | .60 | | .40 | .30 | .29 |
| 10. | AzAs 4 | .28 | .28 | .23 | .32 | .30 | .23 | .48 | .41 | .40 | | .56 | .52 |
| 11. | AzAs 5 | .29 | .28 | .22 | .39 | .32 | .30 | .37 | .35 | .30 | .56 | | .51 |
| 12. | AzAs 6 | .21 | .22 | .13 | .25 | .21 | .19 | .42 | .27 | .29 | .52 | .51 | |

TABLE 12. Comparison of Models (Relative Tolerance: $1 \times 10^{-10}$)

| Sample | Model | DF | $\chi^2$ | p | CFI | TLI | RMSEA Est. | RMSEA Low | RMSEA High |
|---|---|---|---|---|---|---|---|---|---|
| Balanced | One factor | 54 | 2199.29 | <.001 | .88 | .85 | .22 | .21 | .23 |
| Balanced | Two factor | 53 | 633.92 | <.001 | .97 | .96 | .12 | .11 | .12 |
| Balanced | Bifactor | 42 | 24.36 | <.001 | .99 | .98 | .08 | .07 | .09 |
| Balanced | Bifactor w. AB | 41 | 172.82 | <.001 | .99 | .99 | .06 | .05 | .07 |
| Balanced | Bifactor w. 2×AB | 39 | 163.20 | <.001 | .99 | .99 | .06 | .05 | .07 |
| Representative | One factor | 54 | 5244.56 | <.001 | .92 | .90 | .23 | .22 | .23 |
| Representative | Two factor | 53 | 1202.30 | <.001 | .98 | .98 | .11 | .10 | .11 |
| Representative | Bifactor | 42 | 615.49 | <.001 | .99 | .99 | .09 | .08 | .09 |
| Representative | Bifactor w. AB | 41 | 218.80 | <.001 | 1.00 | 1.00 | .05 | .04 | .05 |
| Representative | Bifactor w. 2×AB | 39 | 158.48 | <.001 | 1.00 | 1.00 | .04 | .03 | .05 |

TABLE 13. Fit Indices for Bifactor Model with Two Acquiescence Bias Factors Across Subsamples by Age and Gender (Relative Tolerance: $1 \times 10^{-7}$)

| Sample, subsample | n | DF | $\chi^2$ | p | CFI | TLI | RMSEA Est. | RMSEA Low | RMSEA High |
|---|---|---|---|---|---|---|---|---|---|
| Balanced, 18–25 | 394 | 39 | 7.21 | .002 | 1.00 | .99 | .05 | .03 | .06 |
| Balanced, 26+ | 412 | 39 | 55.82 | .039 | 1.00 | 1.00 | .03 | .01 | .05 |
| Balanced, Female | 407 | 39 | 66.94 | .004 | 1.00 | .99 | .04 | .02 | .06 |
| Balanced, Male | 395 | 39 | 52.82 | .069 | 1.00 | 1.00 | .03 | .00 | .05 |
| Representative, 18–25 | 144 | 39 | 35.53 | .629 | 1.00 | 1.00 | .00 | .00 | .05 |
| Representative, 26+ | 1709 | 39 | 14.36 | <.001 | 1.00 | 1.00 | .04 | .03 | .05 |
| Representative, Female | 1095 | 39 | 138.58 | <.001 | 1.00 | 1.00 | .05 | .04 | .06 |
| Representative, Male | 758 | 39 | 77.66 | <.001 | 1.00 | 1.00 | .04 | .02 | .05 |

level of relative tolerance, for the reasons discussed above).

As table 14 shows, the bifactor model with two acquiescence bias factors was found to exhibit both metric and scalar invariance across age groups when fitted both to the Balanced and to the Representative samples, and to exhibit both metric and scalar invariance across gender groups in the Representative sample (note that the model was again fitted at a more relaxed level of relative tolerance). In the Balanced sample, metric invariance as measured on change in RMSEA was not achieved with regard to gender (with regard to change in CFI, the threshold was met for both metric and scalar invariance). Removing one item at a time from the model and recalculating invariances across gender revealed that full metric and scalar invariance could be restored through removal of AzAs 4 (see table 15). However, as noted above, this appears to be a characteristic of the Balanced sample rather than a characteristic of the model or the scale: when the same model was fitted to the Representative sample, no items were noninvariant.

Model parameters for the two samples with ninety-five-percent confidence intervals (two-tailed) are presented in tables 16 and 17, excluding loadings for the acquiescence bias factors, which were fixed at 1. (Note that negatively keyed items were reverse-coded before analysis, so all loadings are positive.) It is noted that the AzAs items loaded less strongly onto the general factor than the JpAs items when the model was fitted to the dataset collected from the Balanced sample, but not when the model was fitted to the dataset collected from the Representative sample.

Finally, standard reliability measures were calculated. Item-scale rank-order correlations for the GeAs scale are presented in table 18, for the JpAs subscale in table 19, and for the AzAs subscale in table 20, across all four samples. Cronbach's alpha and Guttman's lambda 6 for GeAs, JpAs, and AzAs are presented in table 13, again for all four samples. Internal reliability was found to be excellent for GeAs in the Representative sample $\alpha = .87$, $\lambda_6 = .90$.[50] It was lower for the two subscales considered individually, $\alpha = .89$, $\lambda_6 = .88$ for JpAs and $\alpha = .82$, $\lambda_6 = .81$ for AzAs, and was consistently higher

**TABLE 14.** Invariances for Bifactor Model with Two Acquiescence Bias Factors (Relative Tolerance: $1 \times 10^{-7}$)

| Sample | Group | Invariance | DF | $\chi^2$ | $p$ | CFI | RMSEA | ΔCFI | ΔRMSEA |
|---|---|---|---|---|---|---|---|---|---|
| Balanced | Age | Configural | 78 | 126.03 | <.001 | 1.00 | .04 | | |
| Balanced | Age | Metric | 99 | 208.87 | <.001 | .99 | .05 | -.003 | .013 |
| Balanced | Age | Scalar | 130 | 191.64 | <.001 | 1.00 | .03 | -.001 | -.005 |
| Balanced | Gender | Configural | 78 | 128.04 | <.001 | 1.00 | .04 | | |
| Balanced | Gender | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.006 | .022 |
| Balanced | Gender | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.005 | .011 |
| Representative | Age | Configural | 78 | 176.12 | <.001 | 1.00 | .04 | | |
| Representative | Age | Metric | 99 | 299.72 | <.001 | 1.00 | .05 | -.002 | .010 |
| Representative | Age | Scalar | 130 | 288.07 | <.001 | 1.00 | .04 | -.001 | -.001 |
| Representative | Gender | Configural | 78 | 216.24 | <.001 | 1.00 | .04 | | |
| Representative | Gender | Metric | 99 | 299.54 | <.001 | 1.00 | .05 | -.001 | .003 |
| Representative | Gender | Scalar | 130 | 343.28 | <.001 | 1.00 | .04 | -.001 | -.002 |

TABLE **15.** Invariances Across Gender for Bifactor Model with Acquiescence Bias Factors in Balanced Sample after Removal of Individual GeAs Items (Relative Tolerance: $1 \times 10^{-7}$)

| Removed | Invariance | DF | $\chi^2$ | $p$ | CFI | RMSEA | ΔCFI | ΔRMSEA |
|---------|-----------|----|----------|-----|-----|-------|------|--------|
| JpAs 1 | Configural | 60 | 96.42 | .002 | 1.00 | .04 | | |
| JpAs 1 | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.006 | .023 |
| JpAs 1 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.005 | .012 |
| JpAs 2 | Configural | 60 | 104.87 | <.001 | 1.00 | .04 | | |
| JpAs 2 | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.005 | .019 |
| JpAs 2 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.004 | .008 |
| JpAs 3 | Configural | 60 | 9.41 | .007 | 1.00 | .04 | | |
| JpAs 3 | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.006 | .026 |
| JpAs 3 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.005 | .015 |
| JpAs 4 | Configural | 60 | 78.95 | .051 | 1.00 | .03 | | |
| JpAs 4 | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.007 | .034 |
| JpAs 4 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.006 | .023 |
| JpAs 5 | Configural | 60 | 89.07 | .009 | 1.00 | .03 | | |
| JpAs 5 | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.007 | .027 |
| JpAs 5 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.006 | .016 |
| JpAs 6 | Configural | 80 | 428.40 | <.001 | .98 | .10 | | |
| JpAs 6 | Metric | 99 | 25.53 | <.001 | .99 | .06 | .011 | -.042 |
| JpAs 6 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | .012 | -.053 |
| AzAs 1 | Configural | 60 | 103.76 | <.001 | 1.00 | .04 | | |
| AzAs 1 | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.006 | .019 |
| AzAs 1 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.005 | .008 |
| AzAs 2 | Configural | 60 | 9.76 | .006 | 1.00 | .04 | | |
| AzAs 2 | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.006 | .026 |
| AzAs 2 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.006 | .015 |
| AzAs 3 | Configural | 60 | 98.80 | .001 | 1.00 | .04 | | |
| AzAs 3 | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.006 | .022 |
| AzAs 3 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.005 | .011 |
| AzAs 4 | Configural | 60 | 145.39 | <.001 | .99 | .06 | | |
| AzAs 4 | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.003 | .002 |
| AzAs 4 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.002 | -.009 |
| AzAs 5 | Configural | 60 | 105.99 | <.001 | 1.00 | .04 | | |
| AzAs 5 | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.006 | .018 |
| AzAs 5 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.005 | .007 |
| AzAs 6 | Configural | 78 | 128.04 | <.001 | 1.00 | .04 | | |
| AzAs 6 | Metric | 99 | 25.53 | <.001 | .99 | .06 | -.006 | .022 |
| AzAs 6 | Scalar | 130 | 265.13 | <.001 | .99 | .05 | -.005 | .011 |

TABLE **16.** Model Parameters for Bifactor Model with Two Acquiescence Bias Factors (not shown), Balanced Sample (Relative Tolerance: $1 \times 10^{-10}$)

| Factor | Indicator | Est. | Low | High | SE | Z | p |
|--------|-----------|------|-----|------|------|-------|-------|
| G | JpAs 1 | 1.00 | 1.00 | 1.00 | 0.00 | | |
| G | JpAs 2 | 0.94 | 0.84 | 1.04 | 0.05 | 18.52 | <.001 |
| G | JpAs 3 | 0.74 | 0.63 | 0.86 | 0.06 | 12.67 | <.001 |
| G | JpAs 4 | 0.83 | 0.73 | 0.94 | 0.05 | 15.71 | <.001 |
| G | JpAs 5 | 0.64 | 0.53 | 0.76 | 0.06 | 10.91 | <.001 |
| G | JpAs 6 | 0.67 | 0.53 | 0.82 | 0.07 | 9.28 | <.001 |
| G | AzAs 1 | 0.57 | 0.42 | 0.72 | 0.07 | 7.60 | <.001 |
| G | AzAs 2 | 0.79 | 0.62 | 0.96 | 0.09 | 8.89 | <.001 |
| G | AzAs 3 | 0.36 | 0.24 | 0.49 | 0.07 | 5.55 | <.001 |
| G | AzAs 4 | 0.33 | 0.20 | 0.46 | 0.07 | 4.85 | <.001 |
| G | AzAs 5 | 0.45 | 0.31 | 0.58 | 0.07 | 6.35 | <.001 |
| G | AzAs 6 | 0.19 | 0.06 | 0.31 | 0.06 | 2.94 | .003 |
| F1 | JpAs 1 | 1.00 | 1.00 | 1.00 | 0.00 | | |
| F1 | JpAs 2 | 1.02 | 0.76 | 1.29 | 0.13 | 7.64 | <.001 |
| F1 | JpAs 3 | 2.05 | 1.31 | 2.80 | 0.38 | 5.41 | <.001 |
| F1 | JpAs 4 | 1.22 | 0.89 | 1.55 | 0.17 | 7.27 | <.001 |
| F1 | JpAs 5 | 1.83 | 1.15 | 2.52 | 0.35 | 5.24 | <.001 |
| F1 | JpAs 6 | 1.97 | 1.21 | 2.74 | 0.39 | 5.05 | <.001 |
| F2 | AzAs 1 | 1.00 | 1.00 | 1.00 | 0.00 | | |
| F2 | AzAs 2 | 0.74 | 0.64 | 0.84 | 0.05 | 14.33 | <.001 |
| F2 | AzAs 3 | 1.00 | 0.89 | 1.10 | 0.06 | 18.08 | <.001 |
| F2 | AzAs 4 | 1.16 | 1.03 | 1.29 | 0.07 | 17.61 | <.001 |
| F2 | AzAs 5 | 0.99 | 0.87 | 1.10 | 0.06 | 16.66 | <.001 |
| F2 | AzAs 6 | 1.30 | 1.13 | 1.47 | 0.09 | 14.85 | <.001 |

TABLE **17.** Model Parameters for Bifactor Model with Two Acquiescence Bias Factors (not shown), Representative Sample (Relative Tolerance: $1 \times 10^{-10}$)

| Factor | Indicator | Est. | Low | High | SE | Z | p |
|--------|-----------|------|-----|------|------|-------|-------|
| G | JpAs 1 | 1.00 | 1.00 | 1.00 | 0.00 | | |
| G | JpAs 2 | 1.01 | 0.93 | 1.08 | 0.04 | 27.75 | <.001 |
| G | JpAs 3 | 0.87 | 0.79 | 0.96 | 0.04 | 20.75 | <.001 |
| G | JpAs 4 | 1.06 | 0.97 | 1.15 | 0.05 | 23.36 | <.001 |
| G | JpAs 5 | 1.00 | 0.91 | 1.09 | 0.05 | 21.79 | <.001 |
| G | JpAs 6 | 0.76 | 0.67 | 0.84 | 0.04 | 17.64 | <.001 |

TABLE **17.** Model Parameters for Bifactor Model with Two Acquiescence Bias Factors (not shown), Representative Sample (Relative Tolerance: $1 \times 10^{-10}$)

| Factor | Indicator | Est. | Low | High | SE | Z | p |
|--------|-----------|------|------|------|------|-------|-------|
| G | AzAs 1 | 1.44 | 1.26 | 1.63 | 0.10 | 15.01 | <.001 |
| G | AzAs 2 | 1.31 | 1.11 | 1.51 | 0.10 | 13.02 | <.001 |
| G | AzAs 3 | 0.92 | 0.71 | 1.12 | 0.10 | 8.81 | <.001 |
| G | AzAs 4 | 1.70 | 1.51 | 1.88 | 0.09 | 17.97 | <.001 |
| G | AzAs 5 | 1.67 | 1.50 | 1.84 | 0.09 | 19.53 | <.001 |
| G | AzAs 6 | 1.46 | 1.30 | 1.62 | 0.08 | 18.08 | <.001 |
| F1 | JpAs 1 | 1.00 | 1.00 | 1.00 | 0.00 | | |
| F1 | JpAs 2 | 0.92 | 0.89 | 0.96 | 0.02 | 55.21 | <.001 |
| F1 | JpAs 3 | 0.95 | 0.91 | 0.99 | 0.02 | 47.15 | <.001 |
| F1 | JpAs 4 | 0.87 | 0.83 | 0.91 | 0.02 | 43.95 | <.001 |
| F1 | JpAs 5 | 0.92 | 0.89 | 0.96 | 0.02 | 46.79 | <.001 |
| F1 | JpAs 6 | 0.88 | 0.83 | 0.92 | 0.02 | 37.83 | <.001 |
| F2 | AzAs 1 | 1.00 | 1.00 | 1.00 | 0.00 | | |
| F2 | AzAs 2 | 1.58 | 1.29 | 1.86 | 0.15 | 10.71 | <.001 |
| F2 | AzAs 3 | 3.74 | 1.54 | 5.94 | 1.12 | 3.33 | .001 |
| F2 | AzAs 4 | 0.85 | 0.60 | 1.10 | 0.13 | 6.70 | <.001 |
| F2 | AzAs 5 | 0.23 | -0.04 | 0.50 | 0.14 | 1.64 | .102 |
| F2 | AzAs 6 | 0.62 | 0.37 | 0.86 | 0.12 | 4.97 | <.001 |

TABLE **18.** Item-scale Rank-order Correlations: GeAs Scale

| Item | Test | Retest | Balanced | Representative |
|------|------|--------|----------|----------------|
| JpAs 1 | .68 | .66 | .65 | .74 |
| JpAs 2 | .62 | .60 | .60 | .71 |
| JpAs 3 | .63 | .61 | .64 | .67 |
| JpAs 4 | .68 | .71 | .58 | .70 |
| JpAs 5 | .65 | .70 | .59 | .72 |
| JpAs 6 | .57 | .59 | .55 | .62 |
| AzAs 1 | .54 | .54 | .61 | .53 |
| AzAs 2 | .49 | .47 | .67 | .58 |
| AzAs 3 | .37 | .38 | .48 | .48 |
| AzAs 4 | .56 | .59 | .53 | .61 |
| AzAs 5 | .58 | .65 | .56 | .61 |
| AzAs 6 | .49 | .50 | .49 | .52 |
| Mean | .57 | .58 | .58 | .62 |

both for GeAs and for its two subscales in the Representative sample than it was in the other samples. For example, in the Balanced sample, reliability for GeAs was $\alpha$ = .83, $\lambda_6$ = .87, reliability for JpAs was $\alpha$ = .87, $\lambda_6$ = .86, and reliability for AzAs was $\alpha$ = .80, $\lambda_6$ = .78. See table 21 for full details.

These represent a very good level of internal reliability for the GeAs scale, and, for the JpAs and AzAs subscales, a level of internal consistency similar to those reported for comparable scales. Small sample studies have found the shortened (thirty-two-item) Levinson and Sanford scale to achieve an internal consistency ranging from $\alpha$ = .91 to $\alpha$ = .98,[51] but that scale contains more than 2.6 times more items (where Cronbach's alpha is biased towards longer scales), and also benefits from a lack of reversed items, which is likely to inflate consistency measures.[52]

**TABLE 19.** Item-scale Rank-order Correlations: JpAs Subscale

| Item | Test | Retest | Balanced | Representative |
|------|------|--------|----------|----------------|
| JpAs 1 | .82 | .81 | .80 | .84 |
| JpAs 2 | .75 | .74 | .75 | .80 |
| JpAs 3 | .73 | .72 | .77 | .77 |
| JpAs 4 | .74 | .78 | .80 | .79 |
| JpAs 5 | .75 | .79 | .75 | .81 |
| JpAs 6 | .77 | .76 | .63 | .75 |
| Mean | .76 | .77 | .75 | .80 |

**TABLE 20.** Item-scale Rank-order Correlations: AzAs Subscale

| Item | Test | Retest | Balanced | Representative |
|------|------|--------|----------|----------------|
| AzAs 1 | .69 | .67 | .70 | .69 |
| AzAs 2 | .63 | .62 | .68 | .71 |
| AzAs 3 | .64 | .62 | .62 | .69 |
| AzAs 4 | .70 | .75 | .72 | .73 |
| AzAs 5 | .62 | .68 | .67 | .68 |
| AzAs 6 | .69 | .69 | .71 | .67 |
| Mean | .66 | .67 | .69 | .69 |

**TABLE 21.** Internal Consistency: GeAs Scale and JpAs and AzAs Subscales

| Scale | Sample | $\alpha$ | $\lambda_6$ |
|-------|--------|----------|-------------|
| GeAs | Test | .83 | .86 |
| GeAs | Retest | .84 | .88 |
| GeAs | Balanced | .83 | .87 |
| GeAs | Representative | .87 | .90 |
| JpAs | Test | .86 | .85 |
| JpAs | Retest | .87 | .86 |
| JpAs | Balanced | .87 | .86 |
| JpAs | Representative | .89 | .88 |
| AzAs | Test | .77 | .76 |
| AzAs | Retest | .78 | .78 |
| AzAs | Balanced | .80 | .78 |
| AzAs | Representative | .82 | .81 |

### Study 4: Discussion

The substantially better fit achieved by the two-factor model than by the one-factor model is evidence that JpAs and AzAs items load onto separate factors and are not interchangeable with one another. However, fit for a two-factor model remains poor. The substantially better fit achieved by the bifactor models as opposed to the two-factor model is evidence that, while JpAs and AzAs items load onto separate factors, those factors express a single underlying latent variable, which we may identify with the generalised antisemitism that the scale was developed to measure. The findings of the confirmatory factor analysis thus appear consistent with the spirit of the IHRA Definition, whose examples presuppose that antisemitism can be expressed in statements about Jews qua Jews and statements about Israel alike, for example, Example 5: "Accusing the Jews as a people, or Israel as a state, of inventing or exaggerating the Holocaust."[53]

The improvements in fit brought by introducing acquiescence bias factors, like the finding of correlation between indices of acquiescence bias in Study 2, support the view that satisficing behaviors play a role in responses to the GeAs scale, which in turn both (a) supports the decision to enforce a balance between positively and negatively keyed items in the GeAs scale in order to filter out the effects of such bias and (b) supports the conjecture that such behaviors may have played an unrecognized role in the findings of other studies using unbalanced questionnaires.

Findings of the invariance analysis indicated that the scale functions equivalently both for males and females and for individuals over and under twenty-five years of age in the Representative sample. In the Balanced sample, the scale functions equivalently for individuals over and under twenty-five years of age, but contains one item, that is, AzAs 4, which is noninvariant for this sample only. Removal of

that item from the scale would result in loss of completeness (as the scale would no longer test for the important idea of Zionist conspiratorial influence on politics beyond Israel), in loss of balance between positively and negatively keyed items (which is crucial to avoid acquiescence bias), and in loss of balance between the JpAs and AzAs subscales. Moreover, there is no theoretical reason to single out that particular item for removal, and it has been found to be invariant in the Representative sample, while indications that members (and especially female members) of the Balanced sample may have been unrepresentative of the wider population have been discussed above. In acknowledgement of Weijters and Baumgartner's argument that changes made to scales in order to improve their psychometric properties may bring benefits only on paper,[54] and of the fact that this particular item has only been found to be noninvariant in the model when fitted to one sample, which was inferior to the other with regard to size and arguably also representativeness, we consider it advisable to retain AzAs 4 in the scale. As Putnick and Bornstein observe, "partially invariant" scales often retain high proportions of noninvariant items,[55] in contrast to this particular scale, which only has a single item, and only in relation to one of the two samples used in the analysis of invariances here. The GeAs scale can therefore be described as invariant across gender with regard to a large representative sample, partially invariant across gender with regard to a smaller sample that may be less representative, and invariant across age groups with regard to both of the aforementioned samples.

## CONCLUSION

The primary contribution of this article is methodological. Both subscales have been found to have excellent convergent-discriminant reliability. Moreover, scores both for the GeAs scale and for each of its subscales considered individually have been found to be highly stable over time, and the factor structure of the GeAs scale

as a whole has been found to be invariant across age in two samples, invariant across gender in the largest sample, and partially invariant across gender in the second-largest sample. Lastly, standard measures of internal reliability were very good. Use of the GeAs scale as a standard instrument could thus bring quantitative research on antisemitism a new level of methodological confidence stemming from balance, scholarly grounding, and extensive validation.

However, this article also makes an important empirical contribution in its own right. The establishment of convergent-discriminant validity supports the view that the JpAs and AzAs subscales do indeed measure attitudes to Jews and Israel, which in turn supports the usual interpretation of the finding that agreement with extreme anti-Israel statements correlates with agreement with statements expressive of classic antisemitic stereotypes about Jews. This study's pre-registered replication of that repeated finding thus provides the most robust confirmation to date of the intuitive point that attitudes to Israel are closely related to attitudes to Jews qua Jews. Moreover, that finding has been shown both to be robust to outliers and to hold true across age and gender (despite the novel finding of weaker correlation among female respondents).

Lastly, the theoretical implications of the factor analysis presented here can be emphasised. It has been argued that the factor structure found to achieve best fit with the data collected for Studies 3 and 4 accords with theoretical expectations derived from scholarship on the "new" antisemitism and reflected in the IHRA Definition—to wit, that there is a latent trait, which may be expressed both in statements of the kind associated with the "old" and in statements of the kind associated with the "new" antisemitism (although, for reasons unexplored in this article, individuals may incline more towards the former or the latter). Indeed, this finding may be offered as a potential explanation for the aforementioned repeat finding of imperfect positive correlation between attitudes towards Jews qua Jews and attitudes towards the Jewish state.

## TECHNICAL NOTE

All calculations and visualizations were carried out using R v. 3.6.3.[56] The following R packages were used for key calculations: pwr v. 1.3–0,[57] which implements Cohen's procedures for power analysis,[58] psych v. 2.0.12 for internal reliability and comparisons of correlations (including Williams tests),[59] lavaan v. 0.6–7 for confirmatory factor analysis,[60] and moments v. 0.15 for calculation of skewness and kurtosis.[61] Visualizations were created using ggplot2 v. 3.3.3 and semPlot v. 1.1.2.[62] Drafts of this article were compiled using knitr v. 1.30,[63] with use of kableExtra v. 1.3.1 for table construction.[64] Bootstrapping of confidence intervals for rank-order correlations was accomplished using boot v. 1.3–24.[65]

## REFERENCES

1 Caroline C. Kaufman et al., "Psychological Research Examining Antisemitism in the United States: A Literature Review," *Antisemitism Studies* 4, no. 2 (2020): 237–269.

2 Pierre-André Taguieff, *Rising from the Muck: The New Anti-Semitism in Europe* (Chicago: Ivan R. Dee, 2004); Robert Wistrich, *The Politics of Ressentiment: Israel, Jews, and the German Media* (Jerusalem: Vidal Sassoon International Center for the Study of Antisemitism, 2004); Walter Laqueur, *The Changing Face of Anti-Semitism: From Ancient Times to the Present Day* (Oxford: Oxford University Press, 2006).

3 "Working Definition of Antisemitism," Holocaust Remembrance, https://www.holocaustremembrance.com/resources/working-definitions-charters/working-definition-antisemitism; see Bernard Harrison and Lesley Klaff, "The IHRA Definition and Its Critics," in *Contending with Antisemitism in a Rapidly Changing Political Climate,* ed. Alvin H. Rosenfeld (Bloomington: Indiana University Press, 2021) for discussion.

4 John J. Ray, "Reviving the Problem of Acquiescent Response Bias," *Journal of Social Psychology* 121 (1983): 81–96.

5 Jacques B. Billiet, "Do Unbalanced Scales Influence the Respondent's Opinion?," *Contributions to Methodology and Statistics* 10 (1995): 67–83.

6 Daniel J. Levinson and R. Nevitt Sanford, "A Scale for the Measurement of Anti-Semitism," *Journal of Psychology* 17 (1944): 339–370.

7 D. Gabrielle Jones-Wiley et al., "A Research Note on the Levinson and Sanford Anti-Semitism Scale," *Perceptual and Motor Skills* 105 (2007): 1026.

8 Charles Y. Glock and Rodney Stark., *Christian Beliefs and Anti-Semitism* (New York: Harper & Row, 1966).

9 Such as Glock and Stark's own, as well as that of Anti-Defamation League, "About the Survey Methodology," ADL, https://global100.adl.org/about/global100.

10 Gertrude J. Selznick and Stephen Steinberg, *The Tenacity of Prejudice: Anti-Semitism in Contemporary America* (New York: Harper & Row, 1969).

11 See Tom W. Smith, "A Review: Actual Trends or Measurement Artifacts? A Review of Three Studies of Anti-Semitism," *Public Opinion Quarterly* 57, no. 3 (1993): 380–393 for subsequent versions.

12 E. H. Kaplan and C. A. Small, "Anti-Israel Sentiment Predicts Anti-Semitism in Europe," *Journal of Conflict Resolution* 50, no. 548 (2006); Steven K. Baum and Masato Nakazawa, "Anti-Semitism versus Anti-Israeli Sentiment," *Journal of Religion and Society* 9 (2007): 1–8; Peter Beattie, "Anti-Semitism and Opposition to Israeli Government Policies: The Roles of Prejudice and Information," *Ethnic and Racial Studies* 40, no. 15 (2017): 2749–2767; L. Daniel Staetsky, *Antisemitism in Contemporary Great Britain: A Study of Attitudes towards Jews and Israel* (London: Institute for Jewish Policy Research, 2017); L. Daniel Staetsky, "The Left, the Right, Christians, Muslims, and Detractors of Israel: Who Is Antisemitic in Great Britain in the Early 21st Century?," *Contemporary Jewry* 40 (2020): 259–292.

13 Daniel Allington, David Hirsh, and Louise Katz, "The Generalized Antisemitism (GeAs) Scale: A Questionnaire Instrument for Measuring Antisemitism as Expressed in Relation Both to Jews and to Israel," *Journal of Contemporary Antisemitism* 5, no. 1 (2022): 37–48.

14  *Annual Antisemitism Barometer: 2015 Full Report,* Campaign against Antisemitism, 2015; *Antisemitism Barometer 2017,* Campaign against Antisemitism, 2017; see Daniel Allington, "Judeophobic Antisemitism Among British Voters, 2016–2020," *Journal of Contemporary Antisemitism* 3, no. 2 (2020): 31–38 for analysis.

15  Daniel Allington and David Hirsh, "The AzAs (Antizionist Antisemitism) Scale: Measuring Antisemitism as Expressed in Relation to Israel and Its Supporters," *Journal of Contemporary Antisemitism* 2, no. 2 (2019): 43–51.

16  Jacob Cohen, *Statistical Power Analysis for the Behavioural Sciences* (Mahwah: Lawrence Erlbaum Associates, 1988), 79.

17  J. J. Arnett, "Emerging Adulthood: A Theory of Development from the Late Teens through the Twenties," *American Psychologist* 55, no. 5 (2000): 469–480.

18  A. Reifman, J. J. Arnett, and M. J. Colwell, "Emerging Adulthood: Theory, Assessment and Application," *Journal of Youth Development* 2, no. 1 (2007): 37–48; L. P. Spear, "The Adolescent Brain and Age-Related Behavioral Manifestations," *Neuroscience & Biobehavioral Reviews* 24, no. 4 (2000): 417–463.

19  Arnett, "Emerging Adulthood."

20  A. Furlong, ed., *Handbook of Youth and Young Adulthood: New Perspectives and Agendas* (New York: Routledge, 2009).

21  C. S. Crandall, A. Eshleman, and L. O'Brien, "Social Norms and the Expression and Suppression of Prejudice: The Struggle for Internalization," *Journal of Personality and Social Psychology* 82 (2002): 359–378; E. L. Paluck, "Peer Pressure against Prejudice: A High School Field Experiment Examining Social Network Change," *Journal of Experimental Social Psychology* 47 (2011): 350–358.

22  M. Hjerm, M. A. Eger, and R. Danell, "Peer Attitudes and the Development of Prejudice in Adolescence," *Socius* 4 (2018): article no. 2378023118763187.

23  See Anthony Wells, "How YouGov Does Ethnic Minority Polling," YouGov, June 26, 2020, https://yougov.co.uk/topics/politics/articles-reports/2020/06/26/how-yougov-does-ethnic-minority-polling.

24  See "Graduates in the UK Labour Market: 2017," Office for National Statistics, November 24, 2017, https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/articles/graduatesintheuklabourmarket/2017, which clarifies that just forty-two percent of British adults aged between twenty-one and sixty-four are educated to the same level, and note that, outside that age range, the percentage can be assumed to be lower, as graduation before the age of twenty-one is unusual and access to higher education has increased over time.

25  "ANES Continuity Guide [1952–2016]," American National Election Studies, https://electionstudies.org/resources/anes-continuity-guide/.

26  Gregory M. Herek, "Heterosexuals' Attitudes toward Bisexual Men and Women in the United States," *Journal of Sex Research* 39, no. 4 (2002): 264–274.

27  John F. Dovidio et al., "Perspective and Prejudice: Antecedents and Mediating Mechanisms," *Personality and Social Psychology Bulletin* 30, no. 12 (2004): 1537–1549.

28  Chris G. Sibley et al., "Prejudice toward Muslims in New Zealand: Insights from the New Zealand Attitudes and Values Study," *New Zealand Journal of Psychology* 49, no. 1 (2020): 48–72.

29  For example, a feeling thermometer question for Israel was included in the 2004 and 2008 iterations of the American National Election study, "ANES Continuity Guide [1952–2016]."

30  Bernard Harrison and Lesley Klaff, "In Defence of the IHRA Definition," *Fathom,* January 2020, https://fathomjournal.org/in-defence-of-the-ihra-definition.

31  Daniel Allington, "Convergent-Discriminant Testing of GeAs Subscales," OSF, April 2021, https://doi.org/10.17605/OSF.IO/VZGQM.

32  *British Social Attitudes* 2008, NatCen, 2008, https://bsa.natcen.ac.uk/media/38990/bsa-26-annotated-questionnaires-2008.pdf.

33  Allington, "Convergent-Discriminant Testing of GeAs Subscales."

34  Robert Wistrich, "Anti-Zionism and Anti-Semitism," *Jewish Political Studies Review* 16, nos. 3–4 (2004): 30.

35  Kaplan and Small, "Anti-Israel Sentiment Predicts Anti-Semitism in Europe"; Baum and Nakazawa, "Anti-Semitism versus Anti-Israeli Sentiment"; Beattie, "Anti-Semitism and Opposition to Israeli Government Policies"; Staetsky, *Antisemitism in Contemporary Great Britain*; Staetsky, "The Left, the Right, Christians, Muslims, and Detractors of Israel."

36  Allington, "Convergent-Discriminant Testing of GeAs Subscales."

37  Ibid.

38  Which is very close to the correlation reported by Staetsky, *Antisemitism in Contemporary Great Britain*.

39  Baum and Nakazawa, "Anti-Semitism versus Anti-Israeli Sentiment," 5.

40  Ray, "Reviving the Problem of Acquiescent Response Bias"; McKee J. McClendon, "Acquiescence and Recency Response-Order Effects in Interview Surveys," *Sociological Methods and Research* 20, no. 1 (1991): 60–103.

41  Beattie, "Anti-Semitism and Opposition to Israeli Government Policies," 2755.

42  Joseph K. Goodman, Cynthia E. Cryder, and Amar Cheema, "Data Collection in a Flat World: The Strengths and Weaknesses of Mechanical Turk Samples," *Journal of Behavioral Decision Making* 26 (2013): 213–224.

43  Jesse Chandler, Pam Mueller, and Gabriele Paolacci, "Nonnaïveté among Amazon Mechanical Turk Workers: Consequences and Solutions for Behavioral Researchers," *Behavior Research Methods* 46, no. 1 (2014): 112–130.

44  D. V. Cicchetti, "Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology," *Psychological Assessment* 6, no. 4 (1994): 286.

45  Victoria Savalei and Carl F. Falk, "Recovering Substantive Factor Loadings in the Presence of Acquiescence Bias: A Comparison of Three Approaches," *Multivariate Behavioral Research* 49, no. 5 (2014): 407–424; Francisco J. Abad et al., "Modeling General, Specific, and Method Variance in Personality Measures: Results for ZKA-PQ and NEO-PI-r," *Assessment* 25, no. 8 (2016): 959–977.

46  McClendon, "Acquiescence and Recency Response-Order Effects in Interview Surveys."

47  For example, Rens van de Schoot, Peter Lugtig, and Joop Hox, "A Checklist for Testing Measurement Invariance," *European Journal of Developmental Psychology* 9, no. 4 (2012): 488.

48  Yan Xia and Yanyun Yang, "RMSEA, CFI, and TLI in Structural Equation Modeling with Ordered Categorical Data: The Story They Tell Depends on the Estimation Methods," *Behavior Research Methods* 51 (2018): 421.

49  F. F. Chen, "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance," *Structural Equation Modeling: A Multidisciplinary Journal* 14, no. 3 (2007): 464–504; G. W. Cheung and R. B. Rensvold, "Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance," *Structural Equation Modeling* 9, no. 2 (2002): 233–255; Herbert W. Marsh, Benjamin Nagengast, and Alexandre J. S. Morin, "Measurement Invariance of Big-Five Factors over the Life Span: ESEM Tests of Gender, Age, Plasticity, Maturity, and La Dolce Vita Effects," *Developmental Psychology* 49, no. 6 (2013): 1194–1218; M. Lytkina Rudnev, "Testing Measurement Invariance for a Second-Order Factor: A Cross-National Test of the Alienation Scale," *Methods, Data, Analyses: A Journal for Quantitative Methods and Survey Methodology* 12, no. 1 (2018): 47–76.

50  Compare Cicchetti, "Guidelines, Criteria, and Rules of Thumb for Evaluating Normed and Standardized Assessment Instruments in Psychology," 286.

51  Jones-Wiley et al., "A Research Note on the Levinson and Sanford Anti-Semitism Scale"; John Askew et al., "Evidence for Validity of the Revised Levinson and Sanford Anti-Semitism Scale," *Psychological Reports* 103 (2008): 604–606.

52  See Bert Weijters and Hans Baumgartner, "Misresponse to Reversed and Negated Items in Surveys: A Review," *Journal of Marketing Research* 49, no. 5 (2012): 737–747.

53  "Working Definition of Antisemitism."

54  Weijters and Baumgartner, "Misresponse to Reversed and Negated Items in Surveys."

55  Diane L. Putnick and Marc H. Bornstein, "Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research," *Developmental Review* 41 (2016): 71–90.

56  *R: A Language and Environment for Statistical Computing* (Vienna: R Foundation for Statistical Computing, 2020), http://lib.stat.cmu.edu/R/CRAN/doc/manuals/r-devel/fullrefman.pdf.

57  Stephane Champely, *pwr: Basic Functions for Power Analysis*, 2020, https://CRAN.R-project.org/package =pwr.

58  Cohen, *Statistical Power Analysis for the Behavioural Sciences*.

59  William Revelle, *psych: Procedures for Psychological, Psychometric, and Personality Research* (Evanston: Northwestern University Press, 2020), https://CRAN.R-project.org/package =psych.

60  See Yves Rosseel, "lavaan: An R Package for Structural Equation Modeling," *Journal of Statistical Software* 48, no. 2 (2012): 1–36.

61  Lukasz Komsta and Frederick Novomestky, *moments: Moments, Cumulants, Skewness, Kurtosis and Related Tests*, 2015, https://CRAN.R-project.org/package =moments.

62  See Hadley Wickham, *Ggplot2: Elegant Graphics for Data Analysis* (New York: Springer, 2016), https://ggplot2.tidyverse.org; Sacha Epskamp, "semPlot: Path Diagrams and Visual Analysis of Various SEM Packages," *Output,* 2019, https://CRAN.R-project.org/package =semPlot.

63  Yihui Xie, *knitr: A General-Purpose Package for Dynamic Report Generation in r,* 2021, https://yihui.org/knitr/; see Yihui Xie, "Knitr: A Comprehensive Tool for Reproducible Research in R," in *Implementing Reproducible Computational Research*, ed. Victoria Stodden, Friedrich Leisch, and Roger D. Peng (London: Chapman & Hall, 2014); Yihui Xie, *Dynamic Documents with R and Knitr*, 2nd ed. (Boca Raton: Chapman & Hall, 2015), https://yihui.org/knitr/.

64  Hao Zhu, *kableExtra: Construct Complex Table with "Kable" and Pipe Syntax,* 2020, https://CRAN.R-project.org/package=kableExtra.

65  Angelo Canty and B. D. Ripley, *boot: Bootstrap r (s-Plus) Functions,* 2019; see A. C. Davison and D. V. Hinkley, *Bootstrap Methods and Their Applications* (Cambridge: Cambridge University Press, 1997).