EXPO-12: Development and Validation of the Exposure to Violent Extremism Scale.

ABSTRACT

OBJECTIVE. This study details the development and validation of the 12-item Exposure to Violent Extremism Scale (EXPO-12). We aimed to undertake a transparent and robust process of scale development to present a tool to facilitate research on the relationship between exposure and violent extremism.

METHOD: First, we generated an initial item pool and evaluated items via expert feedback (*n* = 6) and a task designed to assess item comprehension (*n* = 13; Study 1). Second, we explored the underlying factor structure with exploratory factor analysis (EFA) and evaluated item characteristics with item response theory (IRT) in a representative sample of the UK population (*n* = 1, 509; Study 2). Finally, we sought to replicate the factor structure proposed by Study 2 via confirmatory factor analysis (CFA) and examined convergent validity with a related construct, violent extremist intentions (*n* = 1, 475; Study 3).

RESULTS: Study 1 resulted in a preliminary pool of 40 items. Study 2 used EFA to establish a four-factor structure consisting of 21 items. IRT further reduced the item pool by evaluating differential item functioning , discrimination, and location parameters, resulting in EXPO-12. Study 3 replicated the factor structure proposed in Study 2 via CFA. EXPO-12 demonstrated good convergent validity with violent extremist intentions.

CONCLUSION: EXPO-12 is presented as a preliminarily validated measure of the concept, alongside it’s limitations. The scale’s main implication is as a tool to facilitate research to unpack the complexity and nuances of the relationship between exposure and violent extremism.

Introduction

Exposure to terrorism-supportive people, places, and settings is theorised to be a key developmental element in pathways to violent extremism (Taylor & Horgan, 2006). Wiktorowicz (2004) highlights that "while grievances create *potential* participants and selective incentives offer inducements, individuals still need exposure to the movement." In other words, exposure is often a precursor to violent extremism, where it is difficult to conceive of someone committing an act of ideological violence without prior exposure to said ideology. Previous research often considers the role of exposure in violent extremism, however no validated psychometric tool to measure the construct yet exists. Hence, we set out to develop a psychometric scale to facilitate further research into the role of exposure in violent extremism.

Research on exposure to violent extremism

Empirical research demonstrates a tentative causal relationship between exposure and extremist attitudes. For instance, a systematic review of research on the links between exposure to online radical content and violent radicalisation identified 10 empirical studies supportive of this position (Hassan et al., 2018). Results suggest that exposure to radical content online may be associated with extremist *attitudes*. Considering motivation and/or action, previous research examining the effects of both online and offline exposure to violent extremism, similarly suggests that exposure may be related to violent political *action*, (Hassan et al., 2018; Gill et al., 2015; Pauwels & Schils, 2016; Pauwels et al., 2014; Perry et al., 2018).

However, much remains to be understood. For instance, many do not experience negative consequences from viewing extremist materials (Gerstenfeld et al., 2003; Keipi et al., 2017). This is perhaps due in part to differences in propensity (e.g., individual-level susceptibilities) and types of exposure (e.g., self-initiated versus inadvertent) (Bouhana, 2019). Further, a recent meta-analysis summarised much of what we currently know about violent extremism and found seemingly small effect sizes for violent media and violence exposure on radical attitudes, however, larger effects sizes for deviant peers – another facet of exposure. This was true too in terms of radical action, where radical peers had a relatively large effect as a risk factor for radical action (Wolfowicz et al., 2020).

Relatedly, more recent research demonstrated that when seeking out exposure to extremist materials online, not all types of media were equally predictive of a violent extremist intentions (Frissen, 2021). Examining the effects of different types of jihadist materials, Frissen (2021) found self-initiated exposure to seemingly extreme radical material, such as beheading videos, to be surprisingly prevalent among a sample of youth. He suggests curiosity may sometimes be a driver of exposure, hence the *motivation* for seeking out exposure may be relevant to its effects. Such research demonstrates the complexity and the nuance of the relationship between exposure and violent extremism and highlights the need for further research. To pursue such knowledge, it is necessary to operationalise exposure for scientific research.

Operationalising exposure to violent extremism

Previous research predominantly operationalises exposure with single items specifically designed for a particular study. For instance, studies often employ single item measures of exposure (Pauwels et al., 2014; Pauwels & Hardyns, 2018; Pauwels & Schils, 2016; Schils & Pauwels, 2016), or ask participants to self-report exposure-related indicators (Clemmow et al., 2020). Others purposefully develop measures for their research (Frissen, 2021). This foundational research provides a much-needed insight into the effects of exposure on violent extremism and underpins the present study. However, related constructs such as radicalisation are often operationalised with validated psychometric scales. Examples include the Activism and Radical Intention Scale (ARIS) (Moskalenko & McCauley, 2009), Belanger et al’s (2019) political violence scale, the Pro-Violence and Illegal Acts in Relation to Extremism Scale (PIARES) (Ozer & Bertelsen, 2018), and the Sympathies for Radicalisation Scale (SyfoR) (Bhui et al., 2014). Scales such as ARIS are implemented widely in research and have been validated either in development or subsequently in further research. Currently, no validated and reliable measure of exposure to violent extremism exists.

The Present Study

Hence the objectives of the present study are to undertake a rigorous process of scale development and validation to determine a measure of exposure to violent extremism for the purposes of scientific research. We aimed to undertake a robust and transparent process to develop a psychometrically sound research tool. To do so, we implemented Carpenter’s (2018) 10 steps for psychometric scale development: (1) Research the intended meaning and breadth of the theoretical concept (2) Determine sampling procedure (3) Examine data quality (4) Verify the factorability of the data (5) Conduct common factor analysis (6) Select factor extraction method (7) Determine number of factors (8) Rotate factors (9) Evaluate items based on a priori criteria (10) Present results, alongside item response theory (IRT), across three studies, and conclude by presenting the 12-item Exposure to Violent Extremism Scale (EXPO-12).

Study 1

Step 1: Research the intended meaning and breadth of the theoretical concept

Conceptual labels and definitions

Study 1’s purpose was to define exposure to violent extremism and to generate an initial item pool. To do so, we reviewed the existing literature on exposure to gain a theoretical understanding of the concept. First, several theoretical models articulate an analytical framework for understanding how exposure relates to violent extremism. Most describe individual-level causal mechanisms that underpin trajectories to violent extremism

(Borum, 2003; Moghaddam, 2005; Neo, 2016; Precht, 2007; Sageman, 2008; Silber & Bhat, 2007; Wiktorowicz, 2004). Some present multi-level conceptual models (McCauley & Moskalenko, 2008; Veldhuis & Staun, 2009) and others present a conceptual framework for understanding involvement in violent extremism (Bouhana, 2019; Taylor & Horgan, 2006). Of these, many describe a process whereby susceptible individuals are exposed to radicalising influences. Reviewing these models informed our theoretical and conceptual understanding of exposure.

Next, we considered existing research on exposure to understand how the construct has been conceptualised and operationalised thus far. Notably, previous studies often conceptualise exposure as *active* (e.g., self-initiated) or *passive* (e.g., inadvertent/accidental) (Pauwels et al., 2014; Pauwels & Schils, 2016; Frissen, 2021). Differentiating exposure in this way is useful as actively seeking out extremist people, settings, or materials is likely to have different outcomes than inadvertently, or passively, being exposed to violent extremism. For instance, research suggests that those who engage in active exposure may be more concerning than those who are passively exposed (Pauwels et al., 2014; Pauwels & Schils, 2016).

Hence, we defined the construct as *exposure to people, places, or settings which promote ideas which characterise extremism as morally legitimate* (Bouhana, 2019). We further differentiated between passive exposure and active exposure, where passive exposure is *inadvertent* and active exposure is *self-initiated.* Hence, we initially conceived of an exposure scale consisting of dimensions relating to active and passive exposure and proceeded to generate an initial item pool as such.

Generating an initial item pool

The process of generating items was threefold. First, we reviewed previously designed items operationalising exposure to violent extremism from the literature described above. Second, we drew items from an existing codebook of risk and protective factors for violent extremism (Gill et al., 2014). The codebook was initially collated to develop a comprehensive database of behavioural indicators for all known instances of lone-actor terrorism in the UK, US, Western Europe, and Australia. The codebook has since undergone several iterations and now spans over 180 variables including behavioural indicators covering the radicalisation, attack preparation, attack, and attack aftermath stages of the event process. For our purposes, we extracted items relevant to exposure, only.

Briefly, the codebook was informed by comprehensive and systematic literature review, consultation with practitioners, evaluation of open- and closed-source data, and expert review. The codebook and database development are described in further detail elsewhere (Gill et al., 2014; Corner et al., 2020). Third, we consulted with experts in the field who risk assessed over 1, 000 cases referred to a specialist Prevent team (an arm of the UK's counterterrorism strategy), to identify exposure-related behaviours in individuals vulnerable to radicalisation and converted these into preliminary items. This resulted in an initial item pool of 40 items – 20 pertaining to active exposure and 20 to passive exposure.

Next, two rounds of feedback were sought, first from an expert panel, and second, via a comprehension task among a sample of the general population.

Feedback from expert panel

Participants. To assess content validity, multiple, independent expert judges should evaluate all items. Typically, expert panels of 5 to 7 judges are recommended (Boateng et al. 2018; Haynes et al. 1995). Hence, we recruited a convenience sample of six experts. Experts were all practitioners working in the Prevent arm of the UK's counterterrorism strategy and included a Prison Counterterrorism Lead (11 years), a Counterterrorism Police Officer (31 years), a Senior Probation Officer (22 years), two Clinical Nurse Specialists (31 and 34 years), and a Consultant Clinical Psychologist (18 years).

Measures and Procedure. Participants were sent an anonymous link to an online survey. The survey required experts to rate each of the 40 items' clarity and relevance along a 5-point Likert scale. If an item was not rated as clear or relevant, experts were asked to elaborate why. The survey concluded with an open-ended question to elicit any further feedback on the overall scale.

Results. All items were rated as relevant and suggested to be retained. However, several items were rated as requiring minor or major modification to improve clarity. For instance, experts highlighted issues with item wording such as the use of terms like 'political' ("*What is meant by political? This may differ from person to person and potentially for those who are vulnerable to extremism/involved in extremism they may not see their actions/involvement etc as political. Perhaps...support violent action for political, moral, or ideological purpose?*"). Upon consideration of the feedback, eight items were removed, and the remaining 32 items were modified.

Item comprehension task

Second, to assess item comprehension, feedback was elicited from a sample of the UK general population.

Participants. Carpenter (2018) suggests sample sizes for pre-tests can range from 5 to 100 participants dependent on the target population, measure, and pre-test design. Given our target population and the nature of the comprehension task, we invited 15 participants from the general population to complete the pre-test task. Participants were recruited via Prolific, an online access panel, and asked to complete a task designed to assess how respondents understood each item. Three participants did not complete the task satisfactorily and so the final sample size was n = 12. Ages ranged from 21 to 59 years old (M = 31.83, SD = 10.88). 50% identified as female and 50% identified as male.

Measures and Procedure. Each of the 40 scale items were presented in turn and participants were instructed to generate a made-up example for each item. For instance, for the item *"Chosen to meet face-to-face with people who support the use of violence to achieve political, religious, or social goals"* one participant responded, "*A child meets someone on a YouTube comments section and decides to meet up with them in a local meet up for white supremacists."*

Results. Reviewing respondents' answers provided an insight into how participants understood each item and allowed us to assess if this aligned with what we intended. As a result, eight items were removed, and four items were modified. The resulting item pool consisted of 24 items (supplementary Table S1). Next, we proceeded to evaluate the factor structure of the preliminary exposure scale.

Study 2

Step 2: Determine sampling procedure.

Large sample sizes are required for exploratory factor analysis to avoid unstable factors (Carpenter, 2018). Recommendations about optimum sample sizes vary, although Comrey and Lee (2013) provide a guide: 50 (very poor), 100 (poor), 200 (fair), 300 (good), 500 (very good), and 1, 000+ (excellent). Alternatively, some suggest that sample size be determined by the ratio of participants to items, with the optimal ratio being at least 20:1 (Carpenter, 2018). The present study recruited a sample of 1, 509 providing a participant to item ratio of ~63:1.

Participants. We surveyed a representative sample of the UK general population (*n* = 1, 509). Participants were again recruited via Prolific. Participants ranged in age from 18 to 78 years old (M = 45.05, SD = 15.59. 51.4% identified as female, 47.6% identified as male, 0.3% identified as non-binary/third gender, 0.1% preferred to self-describe, and 0.3%

declined to identify their gender. 85.3% stated their ethnicity as White, 0.07% as Asian, 0.03% as Black, and 0.01% as Other.

Procedure. The survey consisted of the 24 preliminary exposure items. A secondary purpose of the survey was to assess different measurement instruments used in research on terrorism and so the items appeared alongside six more established measures of related constructs, although the data were not analysed here. Exposure scale items were all measured on a 7-point Likert scale from ‘Never’ to ‘Every day.’

Step 3. Examine data quality

Data quality was assessed primarily by reviewing attention check failure. There was no missing data, and the final sample size was n = 1, 509.

Step 4. Verify the factorability of the data

First, we verified the factorability of the data by assessing Bartlett’s K-squared (p <.05 is acceptable) and Kaiser-Meyer Olkin factor adequacy (KMO >= .06 is acceptable). These are measures of common variance and multivariate normality and should be assessed before conducting factor analysis. If values are acceptable, we can assume we have verified that the data are suitable for factor analysis. Bartlett’s K-squared (741) = 38840.72, p < .001, and KMO = .97 and so we proceeded to the next step.

Step 5 - 7. Conduct factor analysis, select factor extraction method, determine the number of factors

The goal of exploratory factor analysis (EFA) is to uncover the structure, or dimensionality from observed data where the relationship between variables is not known. EFA applying the principal axis factoring estimation method was performed. Principal axis factoring is preferable when the assumptions of normality are violated, as in the present case. As this type of factor analysis is exploratory, the factor structure of our data is unknown. There are several ways to guide to decision-making about how many factors to consider. In the present study we did so by conducting parallel analysis and examining of the scree plot (supplementary Figure S1). Parallel analysis suggested a six-factor solution may be suitable and examining the scree plot suggested four-to-six factor solutions should be considered.

Step 8. Rotate factors

Determining a rotation technique largely depends on whether factors are theorised to be corelated or uncorrelated. Factors relating to active and passive exposure were presumed to be correlated, based on previous theory and research. Hence, promax, an oblique rotation technique suitable when factors a thought to be correlated was chosen.

Step 9-10: Evaluate Items based on a Priori Criteria and Present the Results

EFA can be a useful tool for uncovering the underlying structure of your data whilst also guiding item-pool reduction. The following criteria were outlined a priori to inform decisions about which items to retain and which to discard. First, we set minimum factor loadings to >.30 (Carpenter, 2018). Items which failed to load significantly onto any factor were removed. Further, items loading significantly onto more than one factor (cross-loading >.30) were removed. Second, for a factor to be retained it had to include at least three items, as bi-factor sub-scales are often (but not always) unstable. Finally, we examined corrected item-total correlations. Corrected item-total correlations are a measure of how coherent an item is with the rest of an item set. A cut-off of >.30 was set, where items demonstrating weak corrected item-total correlations (<.30) were removed. Items were evaluated and removed one by one.

We began by first considering a six-factor solution. This included a factor consisting of only two items and so we examined the five-factor alternative before considering the two items for deletion. A five-factor solution also included a factor consisting of only two items, however a four-factor solution not only aligned better with our theoretical understanding of the construct, but all factors consisted of at least three items. Therefore, we proceeded with a four-factor solution.

Inspecting the factor loadings, two items did not significantly load onto any of the factors and so were removed. One item demonstrated cross-loading and so was also removed. The remaining items loaded onto four factors, two factors relating to active exposure and two relating to passive exposure. All items made theoretical sense however some redundancy was observed. Rather than remove items somewhat subjectively at this stage, and before committing to labelling the factors, the decision was made to proceed to the next step of our scale development process to evaluate items more systematically.

Table 1 displays the preliminary factor solution for the exposure scale. Corrected item-total correlations ranged between 0.51 and .70 where values of .30 - .70 are considered acceptable (de Vaus, 2004), hence all items were retained. Next, we sought to examine item-level characteristics by applying item-response theory (IRT) to further refine the exposure scale.

(INSERT TABLE 1 ABOUT HERE)

Item-response theory (IRT)

IRT emerged as an alternative to classical measurement theory as a method for scale or test development. Although there are several differences between the two approaches, there are three main distinctions: 1) IRT places emphasis on individual items, rather than the scale as a whole, 2) IRT identifies how items assess different levels of the attribute being measured, 3) IRT visualises item and scale characteristics which lends itself to interpretation (DeVellis & Thorpe, 2021).

IRT assumptions

There are several assumptions of IRT which must be met. The first is unidimensionality. In the previous section, EFA determined a four-factor solution. Hence the assumption of unidimensionality is violated. In these instances, each factor or sub-scale must be treated as a separate scale where IRT is performed on each sub-scale in turn (ibid). The second assumption is local independence. Local independence assumes that items are not related to one another. Local independence can be assessed by examining the Q3 statistic where correlations >.20 indicate local dependence. Items which demonstrate local dependence should be modified or considered for deletion. The final assumption is item invariance. That is, items do not perform differently for people or groups with different characteristics (i.e., gender, age) who demonstrate the same level of the underlying latent trait. Item invariance can be assessed by evaluating differential item functioning (DIF). Items flagged for DIF should be modified or considered for deletion.

Differential item functioning (DIF)

Before fitting IRT models to the four sub-scales, we examined each for DIF. Items flagged for DIF were removed. We assessed DIF with respect to gender (male, female) and age (<50 years, 50+ years). For Factor 1, one item (act_10) was flagged for DIF in terms of age, meaning that the item functioned differently for people < 50 years old versus those 50+ years old (supplementary Figure S2). In terms of gender, a further item was flagged for DIF (act_12; supplementary Figure S3). Both items were therefore removed.

For Factor 2, no items were flagged for DIF for age or gender. For Factor 3, one item was flagged for DIF for age (pass_9; supplementary Figure S4) and one for gender (pass_11; supplementary Figure S5). For Factor 4, no items were flagged for DIF for age or gender. Next, we fit graded-response models (GRMs) to each of the reduced sub-scales to evaluate item-level attributes. The GRM is the recommended model for conducting IRT with ordered polytomous data, as we have here.

We assessed item fit by examining the index $S\text{-}X^2$ where fit is acceptable when p.$S\text{-}X^2$ >.05 (non-significant) and the Root Mean Square Error of Approximation (RMSEA.$S\text{-}X^2$) <.06. Items which do not meet these cut-offs demonstrate poor fit to the model. However, it should be noted that $S\text{-}X^2$ is sensitive to large sample sizes as we have here, and so should be considered alongside RMSEA.S_X^2. Further, we examined several item-level parameters to assess item-level characteristics. GRM estimates two item parameters, a slope parameter (*a*) and a location parameter (*b*). Slope parameters assess an item's ability to differentiate between different levels of the latent trait. In general, the larger *a*, the more information the item contributes to the scale. Baker and Kim (2017) summarise item discrimination (*a*) as, 0 = No ability; .01 to .04 = Very low; .35 to .64 = Low; .65 to 1.34 = Moderate; 1.35 to 1.69 = High; >1.70 = Very high.

The location parameter (*b*) represents the level of the latent trait at which a respondent has a 50% probability of endorsing a response category. *b* is estimated for each of the response categories. For instance, b_1 represents the level of the latent trait at which a respondent has a 50% chance of endorsing the response category 'Once or twice.' The *b* parameter also provides an estimate of how much information an item provides. For instance, location parameters that are further apart suggest the item provides information across a wide range of values of the latent trait. Both *a* and *b* can be visualised as item information curves. Item information curves are helpful as a graphical representation of how much information an item provides. The height of the curve corresponds to *a* and the width of the curve corresponds to *b*. For each factor, we evaluated both the slope and location parameters, and inspected each item's information curve. At this stage we also inspected the sub-scales for theoretical convergence. Items were removed one by one and GRM models re-fitted until an optimal solution was reached.

Factor 1

All items demonstrated acceptable fit (p.$S\text{-}X^2$ > .05 and RMSEA.S_X^2 < .06) and local independence was established as Q3 correlations were <.20 (Q3 correlations for all items can be found in supplementary Table S2). Hence, we proceeded to examine the estimated parameters summarised by the item information curves. One item (act_2) was removed as it contributed the least amount of information to the sub-scale. Another way to use item information curves curves in item reduction is to look for redundancy (although some redundancy in a scale is often warranted). If two items demonstrate almost identical item information curves curves, it suggests that the items are contributing the same information towards the scale. When further examining item wording, if items are similarly phrased, this may indicate redundancy. Item act_9 was removed as it captured information cover by item act_6, both in terms of its parameters and item wording. Finally, item act_19 was removed as it was assessed as being less relevant, theoretically, to the newly refined sub-scale.

Three items were retained for the final Factor 1 sub-scale. Final item-level statistics and parameters for all four factors are detailed in full in supplementary Table S3. All items demonstrate good discrimination, as indicated by the high discrimination parameters (*a*) and provide information across a wide range of values of theta (*b*). Figure 1 displays the item information curves and Figure 2 displays the category characteristic curves, or response category curves, for the final sub-scale items. Category characteristic curves for polytomous data depict the categorical response curves for each item. Response curves visualise the probability of a person selecting a particular response given a trait level. For instance, in Figure 2 for item act_6, the pink line (P2) depicts the response curve for response category ‘One or twice.’ The peak of this curve corresponds to the level of the latent trait where a respondent is most likely to select this category. If response categories are functioning as intended, as the level of the latent trait increases, so should the probability of an individual endorsing the next response category. In this respect, we would expect to see sequential peaks from the lowest response category to the highest. For Factor 1, all three items demonstrate this pattern and so we conclude that the response categories are performing as intended.

(INSERT FIGURE 1 ABOUT HERE)

In a perfect solution, we should see unique and distinct peaks for each response category, where each category is the most likely to be selected given a certain value of the latent trait. However, this is not always the case in Figure 2. For instance, for item act_6 of Factor 1, the peak of the response curve for P4, the red line, ('Once or twice a year') is subsumed by the response curves either side of it (P3, P5). Therefore, for this item, respondents are never most likely to select the response 'Once or twice a year.' Issues such as these can suggest your response categories may be redundant. For instance, here, it may be worth considering adopting a 5-point Likert scale in preference to a 7-point Likert scale.

(INSERT FIGURE 2 ABOUT HERE)

Factor 2

As with Factor 1, we examined item information curves, model item-level statistics, and theoretical convergence to reduce the preliminary pool of items. One item was removed as it did not demonstrate good fit (act_17). All other items fit the model well and local independence was established. Items pass_14 and pass_20 demonstrated similar item information curves and were phrased similarly, therefore we removed the worse performing item, pass_20. Three items remained and a final GRM model was fitted. For one item (act_15) p.S_X^2 was significant after correcting for multiple comparisons. However, RMSEA.S_X^2 is < .06 and so the item was retained. All other items demonstrate good model fit.

Next, we examined the GRM parameters, item information (Figure 1) and category characteristic curves (Figure 2) for the remaining items. The most discriminating item is act_14 however items act_1 and act_15 contribute information over a wider range of values of the latent trait (observed by the 'wideness' of the curve) and hence the items perform well together as a sub-scale. However, examining the category characteristic curves some response categories appear null as each response curve does not display a unique peak. Again, a 5-point Likert scale may be more appropriate for this sub-scale.

Factor 3

All items fit the model well and local independence was established. All items were retained after evaluation of the item level characteristics. The final item information and category characteristic curves are displayed in Figure 1 and 2, respectively. Again, all items demonstrate satisfactory item-level characteristics, however as with the previous the sub-scales, some of the response categories are null for some items.

Factor 4

All items fit the model well and local independence was established. All items were retained after evaluation of item level characteristics and inspection of the item characteristic curves. As previously, the items were chosen to maximise discrimination and the values of the latent trait covered, to optimise the sub-scale. The category characteristic curves for Factor 4 again suggest a 5-point Likert scale may be preferred. We provide the frequencies of response categories across Study 2 and Study 3 in supplementary Table S4 for interested readers.

Finally, we examined the information function and conditional reliability curves for each sub-scale. The scale information curves are the sum of the item information curves and provide a summary of how well the sub-scales provide information about the latent trait – this is described by the solid blue line in Figure 3. The dashed red line represents the conditional standard errors – in other words, estimated score precision. Smaller values represent more precise estimates. For instance, for Factor 1, the scale information curve suggests the overall scale provides the most information in the range of +1 to +4. As the conditional standard errors mirrors the scale information curve, estimated score precision can be said to be best in the range +1 to +4.

(INSERT FIGURE 3 ABOUT HERE)

Therefore, the Factor 1 sub-scale provides good and reliable information about the latent trait from low to relatively high values of theta (the latent trait). In the very top range of theta, the sub-scale may not perform as well. For Factor 2, the scale information and conditional reliability curves suggest the overall scale provides the most information and is most reliable in the range of +1.5 to +4. For Factor 3, the overall scale provides the most information and is most reliable in the range of 0 to +4, and for Factor 4, the overall scale provides the most information and is most reliable in the range of +0.5 to +4. In sum, the factors provide good information and reliable estimated score precision over a wide range of theta, however less so at the extreme top-end of the scale.

Concluding Study 2, we arrived at the Exposure to Violent Extremism Scale (EXPO-12), consisting of four factors, each including three items ($M = 1.30$, $SD = .189$; Table 2). Two sub-scales relate to active exposure and two relate to passive exposure. Sub-scales were named by researchers after examining the items and identifying overarching 'themes.' Factor 1 "Active seeking" includes items related to self-initiated seeking out exposure to violent extremism by various means, including online, for people and groups who support violent extremism, and for places and/or settings conducive to violent extremism. Factor 2, "Active action" relates to action-oriented exposure to violent extremism, including meeting face-to-face with extremists, attending protests, and searching for information about bombs and/or weapons. In terms of passive exposure, Factor 3, "Passive online" includes items related to inadvertent exposure to violent extremism online, such as on social media. Factor 4, "Passive offline", includes items related to inadvertent exposure to violent extremism, offline, or in 'real' life, such as in your community.

We further examined several measures which assess scale homogeneity and item-level consistency. First, we calculated corrected item-total correlations for the final EXPO-12 scale (Table 2). Acceptable corrected item-total correlations are .30 - .70. Corrected item-total correlations for the EXPO-12 scale ranged from values ranged from .491 to .698 (Table 2). We also examined the inter-item correlation matrix where acceptable cut-offs range between .20 - .80 (Cohen & Swerdlik, 2005). EXPO-12 inter-item correlations ranged between .21 and .60 and the average inter-item correlation was .40 (supplementary Table S5).

(INSERT TABLE 2 ABOUT HERE)

Next, we calculated McDonald's ω as a measure of internal consistency. Acceptable values are > .70. All sub-scales demonstrated acceptable to good internal consistency: Active seeking (McDonald's ω = .790, M =1.22, SD = .0280), Active action (McDonald's ω = .745, M =1.12, SD = .0064), Passive online (McDonald's ω = .841, M =1.57, SD = .0740). Passive offline (McDonald's ω = .870, M =1.40, SD = .0156) and EXPO-12 overall, demonstrates excellent internal consistency (McDonald's ω = .891, M = 1.33, SD = .199).

Study 3

The purpose of Study 3 was to replicate the factor structure identified in Study 2 via confirmatory factor analysis (CFA) and to assess convergent validity. Convergent validity assesses the relationship between a scale, here EXPO-12, and other constructs theorised to be related to it. In our case, we assessed EXPO-12's convergent validity with a measure of violent extremist intentions, RIS (Moskalenko & McCauley, 2011). First, we attempted to replicate the factor structure determined in Study 2.

Method

Participants

The final sample was also recruited via Prolific. As described in Study 2, there is little consensus on calculating minimum sample size for factor analysis however there are some general guidelines. We refer again to Comrey and Lee (2013) who suggest the following: 50 (very poor), 100 (poor), 200 (fair), 300 (good), 500 (very good), and 1, 000+ (excellent). Hence, we collected a sample of n = 1, 475 as large sample sizes are necessary to achieve meaningful results from CFA. Participants were selected randomly but were all current residents of the UK. In terms of gender, 68.8% identified as male, 30.2% identified as female, and 1.0% identified as non-binary/third gender. Ages ranged from 18 to 60 years old where the mean age was 32.81 (SD = 8.60).

Procedure

CFA was performed in R using the 'lavaan' (Rosseel et al., 2020) and 'SemTools' packages (Jorgensen et al., 2016). We agreed a priori that model fit was accepted if the χ^2/df ratio was < 3 (Byrne, 2001), the Comparative Fit Index (CFI) ≥ .90, Tucker Lewis index (TLI) ≥ .90, Root Mean Square Error of Approximation (RMSEA) ≤ .08, Standardized Root Mean Square Residual (SRMR) ≤ .08 (Hu & Bentler, 1999). We applied a robust estimator as the data displayed a skewed distribution, violating the normality assumption. As such, we conducted a maximum likelihood estimation with robust standard errors and a Satorra-Bentler scaled test statistic (Rosseel, 2020).

Results

The four-factor model demonstrated good fit and was accepted: χ^2 (48) = 111.568, p < .001, χ^2/df ratio = 2.32; CFI_{Robust} = .970, TLI_{Robust} = .958, $RMSEA_{Robust}$ = .054; SRMR = .043. Factor loadings were significant ranging form .401 - .839 (Figure 4). A unidimensional model was also examined to ensure the proposed model was the best fitting. The unidimensional model demonstrated poor fit where χ^2(54) = 598.054, p < .001, χ^2/df ratio = 11.08; CFI_{Robust} = .620, TLI_{Robust} = .535, $RMSEA_{Robust}$ = .083; SRMR = .125 providing further evidence for the suitability of a four-factor structure for EXPO-12.

(INSERT FIGURE 4 ABOUT HERE)

We further examined scale reliability and internal consistency. As in Study 2, all factors demonstrated acceptable to good reliability: Active seeking (McDonald's ω = .740, M =1.22, SD = .118), Active action (McDonald's ω = .792, M =1.18, SD = .034), Passive online (McDonald's ω = .794, M =1.73, SD = .106). Passive offline (McDonald's ω = .838, M =1.27, SD = .060) and EXPO-12 overall, demonstrates excellent internal consistency (McDonald's ω = .859, M = 1.34, SD = .256).

Convergent validity assesses the extent to which a construct relates to other measures of theoretically related concepts. As previously described, exposure to violent extremism has been suggested to be related to violent extremist attitudes, beliefs, and intentions (Pauwels et al. 2014). Hence, to further assess construct validity, we examined the extent to which EXPO-12 related to a measure of violent extremist intentions (RIS). EXPO-12 was found to be significantly positively correlated with violent extremist intentions ($r = .33$, $p < .001$).

Discussion

The objective of the present study was to undertake a robust and transparent process of scale development to construct a psychometrically validated measure of exposure to violent extremism. The rationale for doing so was driven by previous research demonstrating a tentative causal link between exposure and violent extremism (Hasan et al., 2018), as well as more recent research questioning the functional relationship between exposure and violent extremism (Frissen, 2021). Findings such as these necessitate further research to unpack the complexity of the differential effects of exposure – an openly available, validated, and reliable measure of exposure to violent extremism may aid some way in facilitating research designed to answer some of these questions (see supplementary Appendix S1 for the final scale in full).

Over three studies we sought to develop and validate EXPO-12. First, we generated an initial item pool from a process of literature review, drawing from a previously designed codebook of risk and protective factors for violent extremism, and in consultation with experts working within the Prevent arm of the UK's counterterrorism strategy. Construct validity was assessed by a panel of experts who provided feedback resulting in the modification of all items. Next, we designed a task to assess item comprehension among a sample of the general population. Items were either modified or removed based on researchers' evaluation of item comprehension.

The next stage applied EFA and assessed items against a priori criteria to arrive at the preliminary four-factor exposure scale. IRT was then applied to evaluate individual item characteristics to generate a parsimonious final scale solution. After examining the performance of each item across the four sub-scales we arrived at EXPO-12. In a third study, we replicated the proposed four-factor structure via CFA. EXPO-12 demonstrated acceptable internal consistency and reliability, suggesting that the scale may be a useful way to operationalise exposure to violent extremism in research.

Limitations

However, it is important to consider the limitations of the present study and what this may mean for any implications. First, considering the sample, Prolific is an online platform which facilitates crowdsourcing samples for research. Whilst Prolific offers a representative sample function, as we employ in Study 2, the sample may be subject to a selection bias in that only those with internet access are able to access the platform. Hence the true representativeness of our sample should be given consideration.

Second, we did not examine the test-retest reliability of EXPO-12 and therefore we were not able to assess the stability of the scale over time. Relatedly, our data are cross-sectional. Future research should collect longitudinal data to test for the predictive validity of EXPO-12 and to conduct further reliability tests.

Last, EXPO-12 will most likely be dependent on the cultural context where the scale is applied and as such, may not be applicable to non-WEIRD (Western, educated, industrialised, rich, and democratic) countries. No universal measure of exposure should be expected, however studies implementing EXPO-12 should think carefully about whether the underlying construct is applicable to their respective context. We encourage future studies to develop and validate instruments to measure exposure in non-WEIRD countries.

Future Research Directions

Understanding the relationship between exposure and violent extremism is a pertinent research question, particularly when seeking to understand how some may come to engage in extremist violence. The EXPO-12 scale presents one way to operationalise exposure to violent extremism in research designed to unpack the complexity of this relationship. For instance, it may be useful to consider the different effects of active versus passive exposure as Pauwels has done previously, by employing the sub-scales of EXPO-12 alongside a measure of violent extremist intentions, such as RIS. It may also be of interest to consider the mediating, if any, effect of exposure on violent extremism by employing EXPO-12 alongside different measures often used as proxies for violent extremism (RIS, ZProso, PAIRES, SyfoR). Alternatively, researchers may examine risk and protective factors for exposure, given previous research suggests a tentative causal link between exposure and violent extremism (Hasan, 2018).

Clinical Implications

Next, we consider the potential clinical implications of research on exposure to violent extremism which may be facilitated in some way by EXPO-12. Understanding the role of exposure in violent extremism is a key concern of relevant stakeholders, including practitioners charged with managing risk among vulnerable populations. Such stakeholders often make use of risk and/or threat assessment tools with differing methodologies (Salman & Gill, 2021). Many of these tools consider, either implicitly or explicitly, exposure to violent extremism as a risk factor for engagement in violent extremism. This is somewhat intuitive as, as mentioned previously, it is difficult to conceive of someone engaging in an act of ideologically motivated violence without prior exposure to said ideology.

Actuarial tools focussed on risk prediction rely on the development of empirically established risk factors. Structured professional judgments require practitioners to gather information on potential risk and protective factors from a structured manual based on empirical evidence. Both approaches require state-of the-art knowledge generated by research to inform guidance and decision-making. EXPO-12 may provide a tool for researchers to help generate this much-needed evidence base to better inform practice. For instance, understanding who may be vulnerable to the effects of exposure to violent extremism may help inform more targeted preventative programming.

Conclusion

Our hope is that EXPO-12 provides a useful measurement tool to help understand the nature of the relationship between exposure and violent extremism. However, it is important to emphasise that EXPO-12 is not designed to be a measure of violent extremism. It is only a measure of *exposure* to violent extremism. Exposure does not always lead to violent extremism - in fact in most instances, it does not. Therefore, the present scale should not be employed as a proxy measure for violent extremism. Rather, it may be a useful way to investigate a possible mechanism via which some *do* come to be involved in violent extremism, moving toward establishing *when and for whom* exposure may be relevant (Bouhana, 2019).

References

Bélanger, J. J., Moyano, M., Muhammad, H., Richardson, L., Lafrenière, M.-A. K., McCaffery, P., … Nociti, N. (2019). Radicalization leading to violence: A test of the 3N model. *Frontiers in Psychiatry, 10*, 42.

Bhui, K., Warfa, N., & Jones, E. (2014). Is violent radicalisation associated with poverty, migration, poor self-reported health and common mental disorders? *PloS One*, *9*(3).

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: a primer. Frontiers in public health, 6, 149.

Byrne, B. M. (2001). Structural equation modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. *International Journal of Testing*, *1*(1), 55–86.

Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures*, *12*(1), 25–

Clemmow, C., Schumann, S., Salman, N. L., & Gill, P. (2020). The Base Rate Study : Developing Base Rates for Risk Factors and Indicators for Engagement in Violent Extremism *. *Journal of Forensic Sciences*. https://doi.org/10.1111/1556-4029.14282

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology press.

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications*. Sage publications.

Frissen, T. (2021). Internet, the great radicalizer? Exploring relationships between seeking for online extremist materials and cognitive radicalization in young adults. *Computers in Human Behavior*, *114*, 106549.

Gerstenfeld, P. B., Grant, D. R., & Chiang, C. (2003). Hate online: A content analysis of extremist Internet sites. *Analyses of Social Issues and Public Policy*, *3*(1), 29–44.

Gill, P., Corner, E., Thornton, A., & Conway, M. (2015). *What are the roles of the Internet in terrorism? Measuring online behaviours of convicted UK terrorists*.

Gill, P., Horgan, J., & Deckert, P. (2014). Bombing Alone: Tracing the Motivations and Antecedent Behaviors of Lone-Actor Terrorists,,. *Journal of Forensic Sciences*, *59*(2), 425–435. https://doi.org/10.1111/1556-4029.12312

Hassan, G., Brouillette-Alarie, S., Alava, S., Frau-Meigs, D., Lavoie, L., Fetiu, A., … Rousseau, C. (2018). Exposure to extremist online content could lead to violent radicalization: A systematic review of empirical evidence. *International Journal of Developmental Science*, *12*(1–2), 71–88.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. Psychological assessment, 7(3), 238.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., Rosseel, Y., Miller, P., Quick, C., … Preacher, K. (2016). *Package 'semTools.'*

Keipi, T., Oksanen, A., Hawdon, J., Näsi, M., & Räsänen, P. (2017). Harm-advocating online content and subjective well-being: A cross-national study of new risks faced by youth. *Journal of Risk Research*, *20*(5), 634–649.

McCauley, C., & Moskalenko, S. (2008). Mechanisms of political radicalization: Pathways toward terrorism. *Terrorism and Political Violence*, *20*(3), 415–433. https://doi.org/10.1080/09546550802073367

Borum, R. (2003). Understanding the Terrorist Mind-Set. *FBI Law Enforcement Bulletin*, (July), 7–10. Retrieved from https://heinonline.org/hol-cgi bin/get_pdf.cgi?handle=hein.journals/fbileb72&section=47

Moghaddam, F. M. (2005, February). The staircase to terrorism a psychological exploration. *American Psychologist*, Vol. 60, pp. 161–169. https://doi.org/10.1037/0003-066X.60.2.161

Moskalenko, S., & McCauley, C. (2009). Measuring political mobilization: The distinction between activism and radicalism. *Terrorism and Political Violence, 21*(2), 239–260.

Moskalenko, S., & McCauley, C. (2011). The psychology of lone-wolf terrorism. *Counselling Psychology Quarterly*, *24*(2), 115–126.

Ozer, S., & Bertelsen, P. (2018). Capturing violent radicalization: Developing and validating scales measuring central aspects of radicalization. *Scandinavian Journal of Psychology*, *59*(6), 653–660.

Pauwels, L. J. R., Brion, F., & De Ruyver, B. (2014). *Explaining and understanding the role of exposure to new social media on violent extremism. An integrative quantitative and qualitative approach (RADIMED): summary*.

Pauwels, L. J. R., & Hardyns, W. (2018). Endorsement for extremism, exposure to extremism via social media and self-reported political/religious aggression. *International Journal of Developmental Science*, *12*(1–2), 51–69.

Pauwels, L. J. R., & Schils, N. (2016). Differential online exposure to extremist content and political violence: Testing the relative strength of social learning and competing perspectives. *Terrorism and Political Violence, 28*(1), 1–29.

Perry, G., Wikström, P.-O. H., & Roman, G. D. (2018). Differentiating right-wing extremism from potential for violent extremism: the role of criminogenic exposure. *International Journal of Developmental Science*, *12*(1–2), 103–113.

Precht, T. (2007). *Home grown terrorism and Islamist radicalisation in Europe - from Conversion to Terrorism*. Retrieved from https://pdfs.semanticscholar.org/b9fe/2777ace4f86dee3ec2cd5a299318f561c9c5.pdf

Rosseel, Y., Jorgensen, T. D., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., … Katsikatsou, M. (2020). *lavaan: Latent Variable Analysis (0.6-3)[Computer software]*.

Sageman, M. (2008, July). A strategy for fighting international Islamist terrorists. *Annals of the American Academy of Political and Social Science*, Vol. 618, pp. 223–231. https://doi.org/10.1177/0002716208317051

Schils, N., & Pauwels, L. J. R. (2016). Political violence and the mediating role of violent extremist propensities. *Journal of Strategic Security*, *9*(2), 70–91.

Taylor, M., & Horgan, J. (2006). A conceptual framework for addressing psychological process in the development of the terrorist. *Terrorism and Political Violence*, Vol. 18, pp. 585–601. https://doi.org/10.1080/09546550600897413

Veldhuis, T., & Staun, J. (2009). Islamist radicalisation: a root cause model. In *Clingendael Security Paper* (Vol. 12). Retrieved from http://www.clingendael.nl/cscp/publications/papers/?id=7740&&type=summary

Wiktorowicz, Q. (2004). Joining the Cause: Al Muhajiroun and Radical Islam. *Roots of Islamic Rad Icalism*, 29. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.689.7010&rep=rep1&type=pdf

Wolfowicz, M., Litmanovitz, Y., Weisburd, D., & Hasisi, B. (2020). A field-wide systematic review and meta-analysis of putative risk and protective factors for radicalization outcomes. *Journal of Quantitative Criminology*, *36*(3), 407–447.

Table 1. Final EFA solution for the exposure scale.

Item No.'s	Item wording	Factor 1	2	3	4	Mean	Std. Deviation	Skewness	Corrected item-total correlation
act_2	Searched for books, magazines, or other types of text which support the use of violence to achieve political, religious, or social goals	0.61				1.192	0.620	3.919	0.506
act_6	Searched for content online like websites, memes, or videos that support the use of violence to achieve political, religious, or social goals	0.89				1.253	0.739	3.658	0.577
act_9	Used the internet to observe online chat between other people who support the use of violence to achieve political, religious, or social goals	0.69				1.189	0.620	4.186	0.634
act_10	Searched for items or memorabilia relating to people or groups who support the use of violence to achieve political, religious, or social goals	0.67				1.307	0.774	2.932	0.639
act_12	Searched for podcasts, songs, or other types of audios which support the use of violence to achieve political, religious, or social goals	0.76				1.262	0.689	3.099	0.612
act_18	Searched for places where people who support the use of violence to achieve political, religious, or social goals spend time	0.74				1.215	0.633	3.786	0.632
act_19	Searched for content made by people who have committed violence to achieve political, religious, or social goals, such as manifestos, or YouTube videos	0.73				1.425	0.886	2.437	0.529
act_21	Searched for people or groups who support violence to achieve political, religious, or social goals	0.60				1.199	0.674	4.483	0.565
act_1	Chosen to meet face-to-face with people who support the use of violence to achieve political, religious, or social goals		0.40			1.178	0.680	4.918	0.540
act_14	Attended protests knowing people who support the use of		0.73			1.050	0.352	9.451	0.694

	violence to achieve political, religious, or social goals would be there							
act_15	Searched for information on how to use weapons or make bombs for violence to achieve political, religious, or social goals	0.74			1.119	0.502	5.797	0.644
act_17	Searched for content about people who have committed violence to achieve political, religious, or social goals	0.98			1.058	0.373	8.465	0.564
act_20	Searched for events or activities to attend which support violence to achieve political, religious, or social goals	0.70			1.072	0.449	8.133	0.638
pass_1	Know of people where you live who support violence to achieve political, religious, or social goals		0.84		1.559	1.096	2.704	0.577
pass_5	Know of activity where you live that supports violence to achieve political, religious, or social goals		0.94		1.400	0.946	3.185	0.595
pass_14	Know of places where you live where activity that supports violence to achieve political, religious, or social goals takes place		0.75		1.248	0.764	4.192	0.700
pass_6	Accidentally came across content which supports violence to achieve political, religious, or social goals online			0.57	1.657	1.042	2.012	0.669
pass_8	Received content online that you didn’t ask for, such as memes or videos which violence to achieve political, religious, or social goals			0.90	1.542	1.023	2.282	0.651
pass_9	Content which supports violence to achieve political, religious, or social goals recommended to you on social media			0.70	1.304	0.809	3.397	0.637
pass_11	Received content online that you didn’t ask for such as images or videos which show acts of violence to achieve political, religious, or social goals			0.83	1.440	0.885	2.488	0.597
pass_15	Came across content online about using violence to achieve political, religious, or social goals while looking for content about something else			0.64	1.519	0.948	2.152	0.652

Table 2. EXPO-12 final scale item statistics

	Item no.'s	Item	Mean	SD	Skew	Item-total correlation
Factor 1	act_6	Searched for content online like websites, memes, or videos that support the use of violence to achieve political, religious, or social goals	1.253	0.739	3.658	0.538
	act_18	Searched for places where people who support the use of violence to achieve political, religious, or social goals spend time	1.215	0.633	3.786	0.608
	act_21	Searched for people or groups who support violence to achieve political, religious, or social goals	1.199	0.674	4.483	0.568
Factor 2	act_1	Chosen to meet face-to-face with people who support the use of violence to achieve political, religious, or social goals	1.178	0.68	4.918	0.491
	act_14	Attended protests knowing people who support the use of violence to achieve political, religious, or social goals would be there	1.05	0.352	9.451	0.494
	act_15	Searched for information on how to use weapons or make bombs for violence to achieve political, religious, or social goals	1.119	0.502	5.797	0.527
Factor 3	pass_6	Accidentally came across content which supports violence to achieve political, religious, or social goals online	1.657	1.042	2.012	0.698
	pass_8	Received content online that you didn't ask for, such as memes or videos which violence to achieve political, religious, or social goals	1.542	1.023	2.282	0.634
	pass_15	Came across content online about using violence to achieve political, religious, or social goals while looking for content about something else	1.519	0.948	2.152	0.635
Factor 4	pass_1	Noticed people where you live who support violence to achieve political, religious, or social goals	1.559	1.096	2.704	0.635
	pass_5	Noticed activity where you live that supports violence to achieve political, religious, or social goals	1.400	0.946	3.185	0.659
	pass_14	Noticed places where you live where activity that supports violence to achieve political, religious, or social goals takes place	1.248	0.764	4.192	0.643

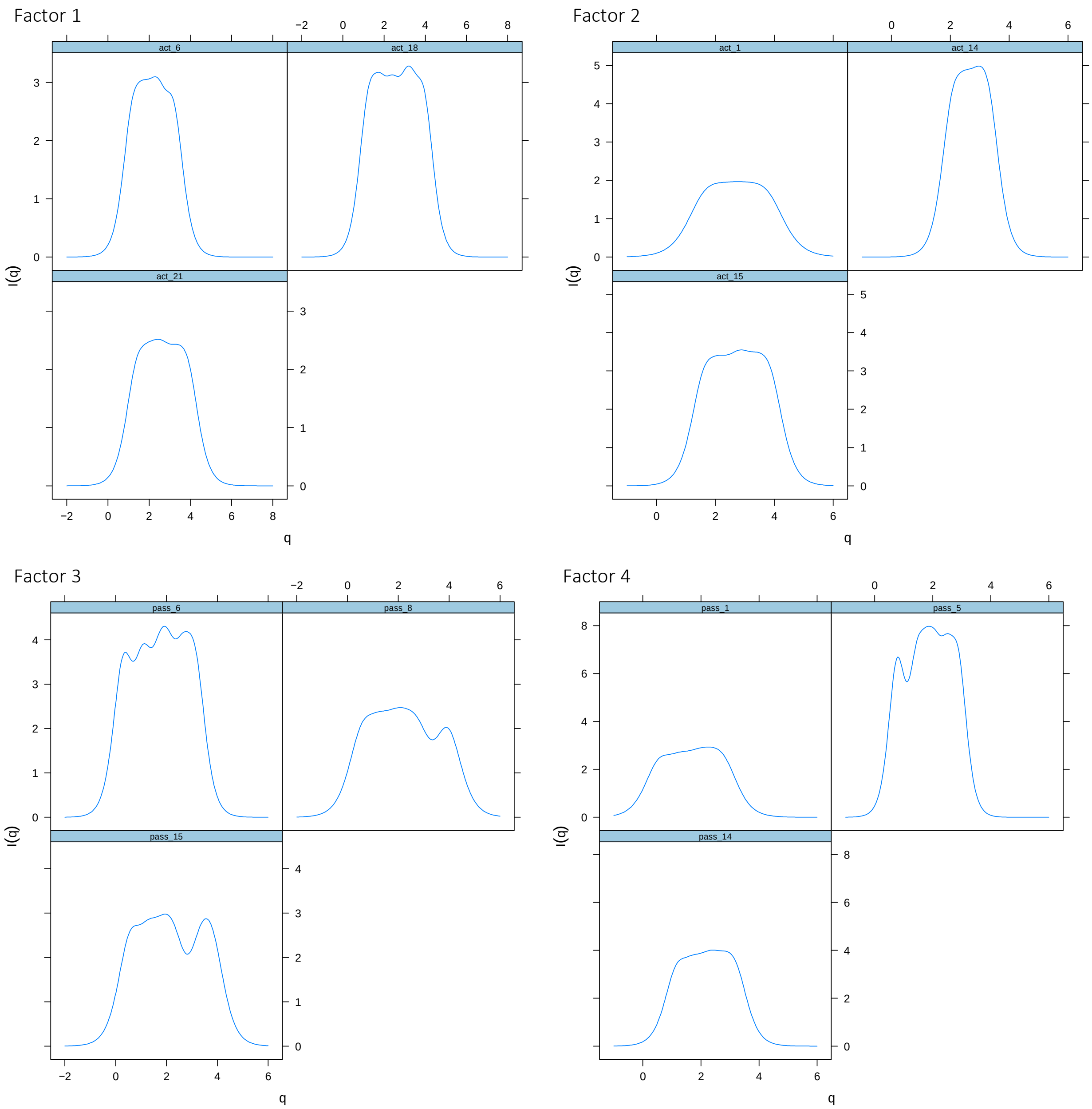

Figure 1. Item Information Curves (IICs) for the four sub-scales of EXPO-12

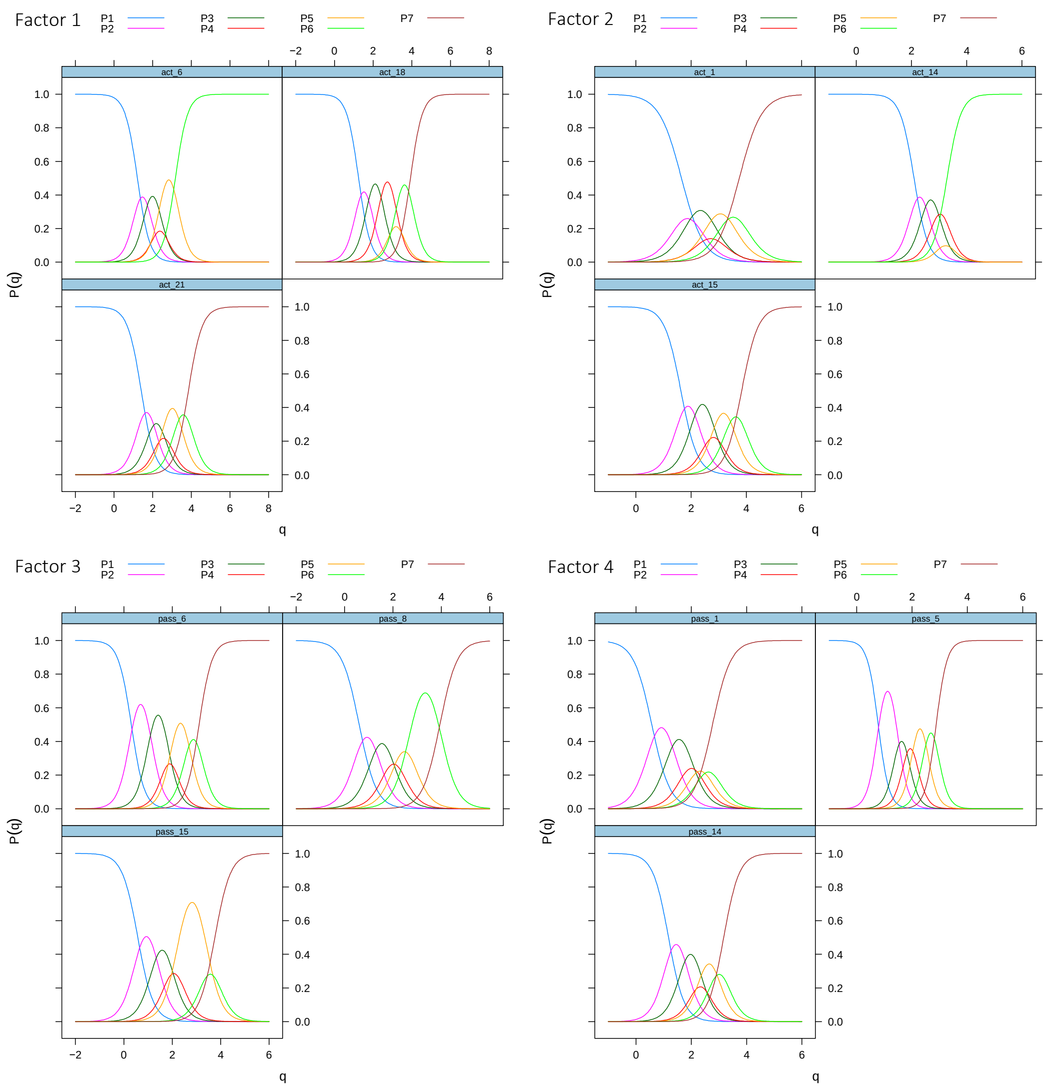

Figure 2. Category Characteristic Curves (CCCs) for the four sub-scales of EXPO-12

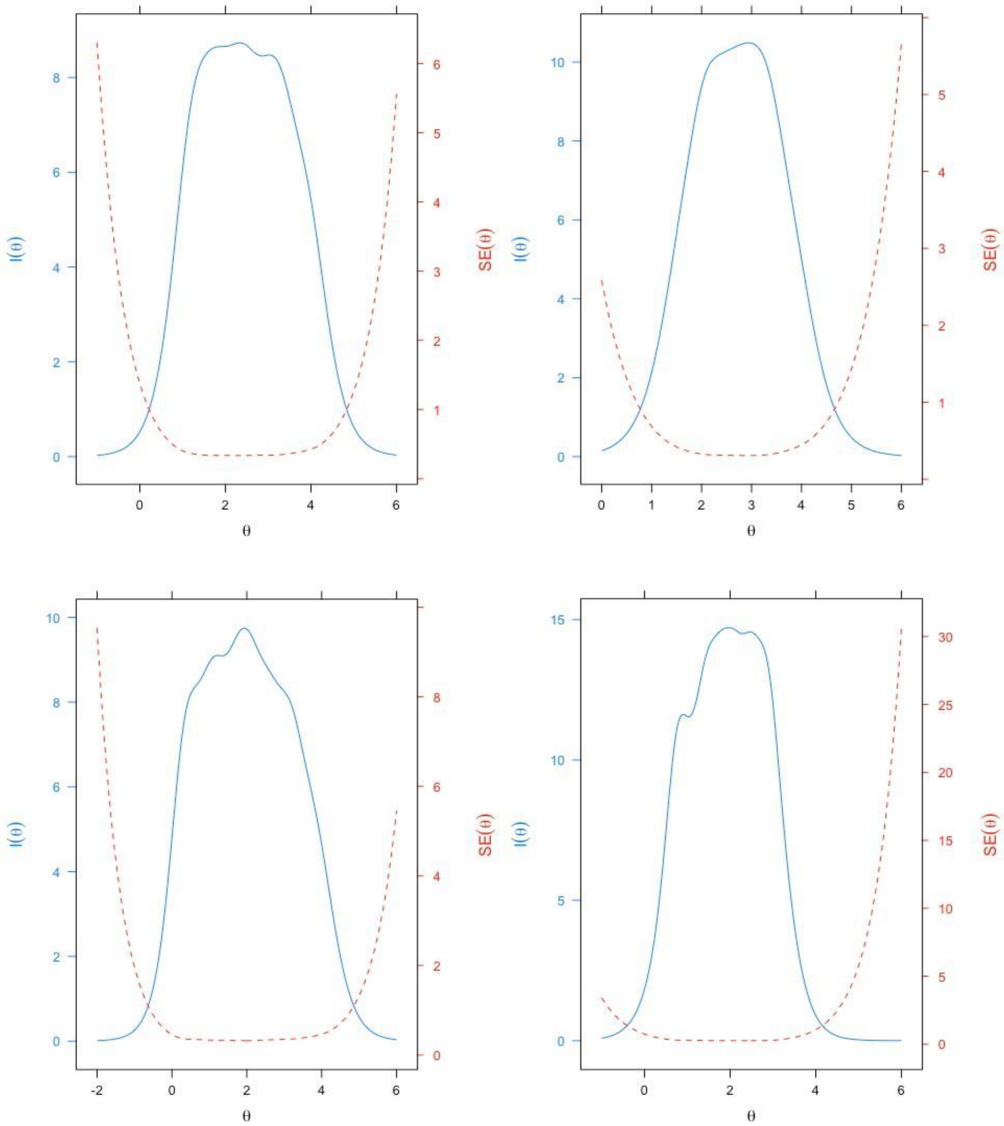

Figure 3. Factor sub-scale information function (solid blue line) and conditional reliability (dashed red line) curves

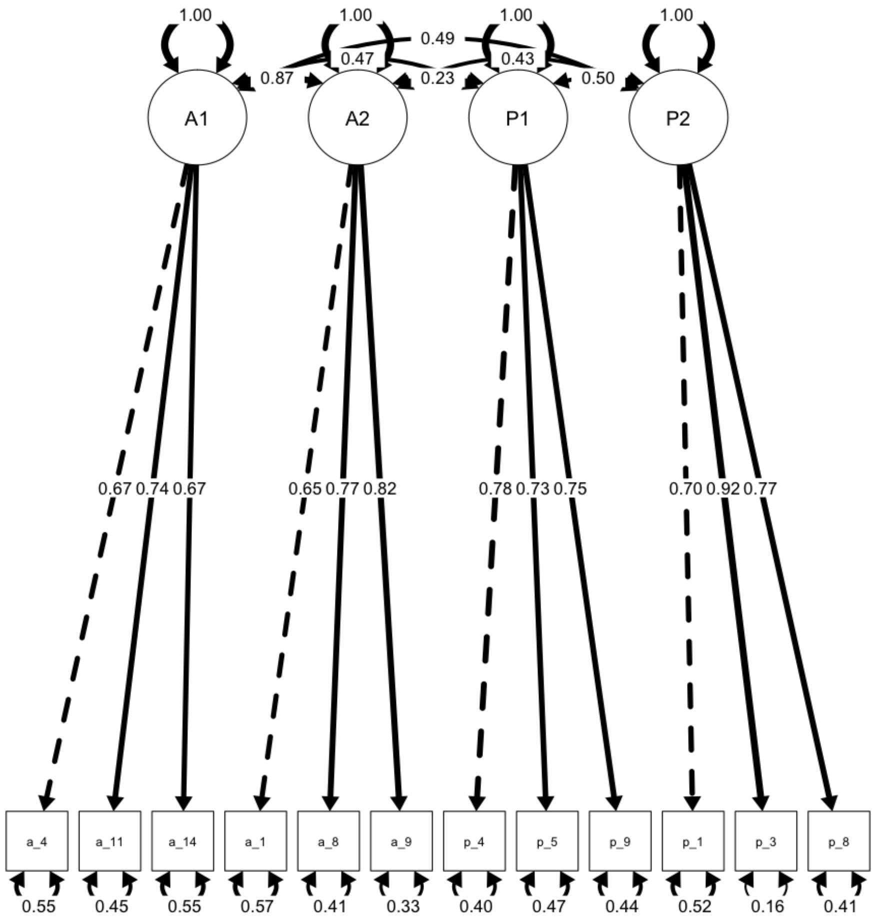

Figure 4. CFA of the four-factor EXPO-12.