

Resource Mobilization under Forced Migration: Development and Validation of the Refugee Resilience Scale (RRS-16)

Öznur Çınar

Bayburt University

Ekrem Cengiz

ekremcengiz@gumushane.edu.tr

Gümüşhane University

Emel Yıldız

Gümüşhane University

Salih Yıldız

Gümüşhane University

Büşra Tosunoğlu Kusan

Gümüşhane University

Research Article

Keywords: forced migration, refugee mental health, resource mobilization, scale development, measurement invariance, psychometrics

Posted Date: September 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10717106/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background.

Resilience in forcibly displaced populations is usually assessed with dispositional instruments developed in general populations, which index the capacity to recover from adversity but not the relational and institutional resources that shape mental health after displacement, although post-migration living difficulties predict distress at least as strongly as trauma exposure. No resilience instrument validated in displaced adults models the capacity to navigate services and rights as a dimension in its own right, although standalone navigation-competency measures have begun to appear for particular service systems.

Methods.

A theory-derived 36-item pool was reduced to 16 items through a three-round modified Delphi review (eight experts), and the retained form was verified in cognitive interviews with 12 displaced adults. After a pilot ($N = 107$) and a Monte Carlo sample size simulation, 583 forcibly displaced Syrian, Afghan and Iraqi adults in Türkiye completed the scale face-to-face; 141 provided a two-week retest. The sample was split into calibration and validation subsamples and analysed by ordinal exploratory and confirmatory factor analysis, exploratory graph analysis, invariance testing, item response theory and network psychometrics.

Results.

Two independent methods recovered three factors: individual adaptive capacity, support seeking and relational resource use, and service and rights navigation. The correlated three-factor model fitted acceptably in the validation subsample ($CFI = .952$; $RMSEA = .064$), and bifactor diagnostics in the full sample did not support a total score (explained common variance = $.496$; hierarchical omega = $.694$). Ordinal alpha was $.884$ for the full scale and $.816$ to $.859$ by factor; two-week stability was $.779$. Scalar invariance held across five grouping variables. All three factors correlated positively with dispositional resilience and self-efficacy and, more weakly and with one non-significant exception, negatively with perceived stress and depressive symptoms, most strongly for adaptive capacity. Longer residence and higher language proficiency raised navigation scores but left adaptive capacity unchanged.

Conclusions.

The RRS-16 provides reliable, structurally replicable and group-comparable measurement of resource mobilization in displaced adults. Its distinctive contribution is the navigation dimension, which general resilience measures do not capture and which varies with settlement. Cross-language equivalence, predictive validity and responsiveness remain to be established.

Background

Displacement is a permanent feature of the global public health landscape. More than one hundred million people are currently displaced by conflict, persecution, and disaster [1]. Elevated depression, anxiety, and post-

traumatic stress persist years after resettlement [2], and a meta-analysis of Syrian refugees in high-income countries shows the elevation is neither transient nor confined to a subgroup [3]. Its determinants extend beyond exposure to violence.

Silove's adaptation and development after persecution and trauma model located refugee distress in the disruption of core psychosocial pillars — safety, attachment, justice, identity, and meaning — rather than in traumatic memory alone [4]. Miller and Rasmussen then showed that daily post-migration stressors explain much of the variance trauma-focused models leave open [5], and a systematic review has since confirmed the association between settlement conditions and common mental disorders across Europe [6]. Among displaced adults in Germany, post-migration difficulties predicted depressive and post-traumatic symptoms more strongly than cumulative trauma, while dispositional resilience buffered distress generally without moderating any of these associations [7].

That result states our problem. Dispositional resilience mattered, but it did not interact with the stressors that mattered most. One reading is that resilience is a weak moderator. The more plausible reading is that the instruments measuring it were never built to capture the resources a post-migration environment withholds — and instruments for those difficulties are themselves few and unevenly developed [8].

The dispositional heritage of resilience measurement

Resilience research took close to the opposite path. Originating in developmental studies of children facing adversity [9, 10], the construct was progressively operationalized as a stable personal characteristic, and the instruments used with displaced populations inherit that framing: the Connor–Davidson Resilience Scale [11, 12], the Brief Resilience Scale [13], and the Resilience Scale [14] were all built in general or clinical populations and index the capacity to recover from adversity. The Resilience Scale for Adults [15] is the partial exception: it models family cohesion and social resources as explicit subscales and so reaches further into relational territory than the others, but it too was developed in a general population and contains no institutional component.

Ungar argued that this misreads resilience wherever the decisive resources are not held by the individual: resilience is the capacity to navigate to and negotiate for culturally meaningful resources, and so a property of the environment as much as of the person [16]. Hobfoll's conservation of resources theory is complementary, treating adaptation as the acquisition, protection, and mobilization of objects, conditions such as legal security, personal characteristics, and energies such as time and information [17, 18].

Two measurement problems follow. The first is population fit. Most resilience instruments used with refugees were developed elsewhere and carry thin evidence in displaced samples [19, 20]; the most widely used has not replicated its five-factor structure, and unidimensional solutions are reported instead [21]. Invariance evidence comes largely from general-population samples [22], and in ethnically heterogeneous refugee samples only partial scalar invariance has been attainable [23]. Comparison across origin and settlement groups is what service planning requires, so this gap is consequential rather than technical.

The second is construct coverage. A meta-analysis found reliable but modest resilience–mental health associations, with heterogeneity traceable to how resilience was operationalized [24] — the pattern expected if available instruments capture only part of what lets displaced people cope. Qualitative work agrees. Emotional coping and social connectedness among Syrian refugees are shaped by the conditions of settlement [25, 26],

and interview research with Iraqi refugees documents networks that contract sharply, support concentrating in the family, and contact with authorities often experienced as humiliating [27] — conditions under which, we would argue, managing a bureaucratic process and managing one's own distress are not separable.

What existing instruments do not cover

Context-specific instruments have begun to address population fit: measures for Syrian refugee adolescents [28] and, from the same social-ecological tradition, an adult instrument organized around individual, relational and contextual resources [29]; Somali adults [30]; household self-reliance indices [32]; and the Refugee Post-Migration Stress Scale [33], the closest precedent here and instructive because it measures stressors rather than the capacity to act on them. Separately, general-population instruments have been evaluated in displaced samples — the Sense of Coherence scale in Eritrean refugees [31], for instance — but revalidating an existing instrument does not add displacement-specific content. Navigation itself has recently begun to be measured in its own right rather than as a facet of resilience: a 35-item healthcare navigation competency scale for refugees resettled in the United States covers system knowledge, insurance, appointments, transport, interpretation and billing [34]. It establishes that the territory can be operationalized, and it also shows the shape of what remains missing: it is confined to a single service system, was developed within one national group, and in its published form is evaluated for content and face validity only, without a field sample, a factor structure, reliability estimates or evidence of comparability across groups. Table 1 sets the coverage of these instruments against four requirements following from the account above. None combines displacement-specific content, an explicit institutional-access dimension, and multi-group invariance evidence in a displaced adult sample.

Table 1
Structural coverage of resilience instruments used with displaced adults

Instrument	Development population	Displacement-specific content	Relational resource dimension	Institutional access dimension	Scalar invariance in displaced adults	Principal limitation
CD-RISC [11, 12]	General and clinical adults	No	Implicit only	No	No; partial only [23]	Five-factor structure does not replicate; unidimensional solutions common [21]
Brief Resilience Scale [13]	General adults	No	No	No	Not reported	Bounce-back capacity only; no resource-mobilization content
Resilience Scale [14]	General adults	No	No	No	Not reported	Original two-factor structure not supported in an immigrant sample [35]; no relational or structural content
Resilience Scale for Adults [15]	General adults	No	Yes; family cohesion and social resources	No	Not reported	Six-factor protective-resource model developed outside displacement; relational content is familial and network-based, with no institutional component
Sense of Coherence scale [36]	General adults	No	No	No	Not reported	Unidimensional 12-item solution fitted better in Eritrean refugees [31]; indexes cognitive orientation, not mobilization
CYRM-12, Arabic [28]	Refugee and host-community adolescents	Yes, for adolescents	Partial	No	Not reported	Adolescent instrument; adult use requires the separate Adult Resilience Measure

Instrument	Development population	Displacement-specific content	Relational resource dimension	Institutional access dimension	Scalar invariance in displaced adults	Principal limitation
Adult Resilience Measure (RRC-ARM) [29]	Adults facing adversity, including conflict-affected and displaced samples	No	Yes	No	Not reported	Social-ecological content is operationalized as belonging, caregiving and religious participation rather than institutional access
SDRS [30]	Somali refugee adults	Yes	Partial	No	Not tested	Combines symptom and resilience subscales; single-community basis
Self-Reliance Index [32]	Displaced households	Yes	No	Partial, economic only	Not applicable	Household economic functioning, not individual capacity
RPMS [33]	Syrian refugee adults	Yes	No	No; measures stressors	Not reported	Measures stressors encountered, not capacity to act on them
HECNAC [34]	Refugee adults resettled in the USA (single national group)	Yes	Implicit only	Yes; healthcare only	None published	A 35-item standalone navigation-competency instrument rather than a resilience dimension; confined to one service system and evaluated for content and face validity only
Refugee Resilience Scale (present study)	Displaced adults in Türkiye	Yes	Yes	Yes	Yes; five groupings, including origin, stay and language	Single language; predictive validity and responsiveness untested

Note. CD-RISC = Connor–Davidson Resilience Scale; CYRM-12 = brief Arabic form of the Child and Youth Resilience Measure; RRC-ARM = Resilience Research Centre Adult Resilience Measure; SDRS = Somali Distress

and Resilience Scale; RPMS = Refugee Post-Migration Stress Scale; HECNAC = HEalthCare NAVigation Competency Scale. The last two rows before the present study are included because they mark the boundary of the construct rather than because they are resilience instruments: one measures the stressors encountered, the other the competency required within a single service system. Coverage is judged against the requirement stated in the text rather than against the developers' own aims; an entry of "No" therefore records the absence of a feature the present construct requires, not a defect of the instrument for its intended purpose. "Implicit only" denotes content from which the dimension could be inferred but which is not structurally differentiated. Invariance entries refer to full scalar invariance established in a forcibly displaced adult sample; "None published" denotes an instrument for which no field-sample psychometric evaluation has yet appeared.

The question is not whether established instruments are inadequate; they measure dispositional recovery capacity well. It is whether forced migration warrants targeted measurement. We argue that it does. Securing legal status, locating care in an unfamiliar system, and completing procedures in a second language are not variants of bouncing back. The resources that decide outcomes are often not held by the individual, so acting depends on whether an accessible institution exists — a contingency hardiness items cannot register. A respondent who bounces back quickly tells us nothing about whether her permit office answers the telephone.

The present instrument and its boundaries

This study develops and evaluates the Refugee Resilience Scale (RRS-16), a measure of adaptive coping and resource mobilization in forced migration. We use *resilience* narrowly and state the boundaries first, because an implicit construct space invites the drift that has made this literature hard to synthesize [37]. Identity continuity, meaning-making, recovery from trauma, cultural belonging, general coping, symptom burden, post-traumatic growth, general self-efficacy, general social support, institutional trust, acculturation, and objective socioeconomic level are all excluded. Two exclusions carry most of the weight: symptom burden, because equating resilience with the absence of symptoms empties the concept of independent content [37, 38]; and institutional trust, because no item asks whether institutions are trustworthy, only whether the respondent can act within them.

What the measure targets is the capacity to mobilize resources across three nested levels, specified a priori.

Individual adaptive capacity (F1) covers problem-focused action, decision-making under uncertainty, plan flexibility, emotional self-regulation, and improvement expectancy, drawing on self-efficacy [39], psychological flexibility [40], hope theory [41], and salutogenic manageability [36]. **Support seeking and relational resource use (F2)** covers perceived availability of support, willingness to seek it, and connection within and beyond one's own community, drawing on support buffering [42] and the bonding–bridging distinction [43]. **Service and rights navigation (F3)** covers information access, service knowledge, procedural competence, and knowing what steps an entitlement requires, drawing on the post-migration stressor literature [5] and the integration framework [44].

The second and third levels belong together because networks contract after displacement and the remaining support is supplemented by contact with authorities [27]. The third is the measure's principal contribution — not because structural access has never been measured, since a dedicated navigation instrument has recently shown that it can be, but because it is brought inside the resilience construct as one of three dimensions rather than left as a background condition or isolated as a separate competency. It complements established instruments rather than replacing them: capability beliefs are domain-specific [39].

One boundary deserves stating, because it is the first objection a careful reader will raise. Knowing where a hospital is does not in itself constitute resilience. The third dimension measures the perceived capacity to act on such knowledge, which is why its items are phrased as capability; items indexing the presence of a service, or the outcome of using one, were eliminated at the first expert round.

Aims and hypotheses

We test four numbered expectations, against which the Results can be read.

H1 (structure). The three-factor structure will be recovered in a calibration subsample by both factor-analytic and network-based methods and will fit acceptably in an independent validation subsample. A single general dimension will not represent the item covariance adequately, and where a bifactor alternative fits better its general factor will be too weak to justify a total score.

H2 (reliability). Internal consistency and composite reliability will be adequate at total and factor levels, and scores stable over two weeks.

H3 (invariance). The measure will function equivalently at configural, metric, and scalar levels across the origin, settlement, and language groups structuring refugee populations in Türkiye, with no practically meaningful differential item functioning.

H4 (validity). All three factors will relate positively to dispositional resilience and self-efficacy and negatively to perceived stress and depressive symptoms, and the associations will be *graded*: individual adaptive capacity strongest throughout, service and rights navigation weakest on distress and more strongly tied to self-efficacy than to dispositional resilience. A directional known-groups hypothesis is pre-specified only for Turkish proficiency.

Methods

Overview and design

Scale development followed a sequential two-phase design grounded in contemporary measurement standards [45–47]. Phase 1 covered construct specification, item generation, expert review, cognitive interviewing, and piloting; Phase 2 covered main-study data collection and psychometric evaluation. Reporting follows the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement [45], the Standards for Educational and Psychological Testing [46], and the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) taxonomy [47]; the study was not preregistered, and criterion validity and responsiveness were not assessed (Additional file 1, Appendix B).

Phase 1: item development

Construct specification and item generation. The construct was specified as the capacity to mobilize resources across the three nested levels defined in the Background – individual adaptive capacity (F1), support seeking and relational resource use (F2), and service and rights navigation (F3) – with items written to index capacity rather than its outcomes. Item generation was deductive [48]: two authors wrote candidate items by iterative consensus against five criteria – construct alignment, the single-idea principle, linguistic simplicity, factor

specificity, and absence of overlap with the adjacent constructs excluded in the Background — producing a 36-item pool of 12 items per factor. Items were written in the first person, referred to no country or ethnic group, avoided clinical vocabulary and absolutes, and were contextualized by a common instruction rather than a repeated stem; F1 items were written so as not to read as general hardiness, F2 items preserved the distinction between perceived availability and active mobilization, and F3 items were written as capability beliefs rather than institutional trust. Three items were flagged a priori as elimination candidates — two approaching the construct boundary, one of them conflating resilience with its outcomes [37, 38], and one likely to cross-load onto F1 — and all three were eliminated at the first expert round. The full construct specification, the initial 36-item pool and the item-level source and risk matrices are documented in Additional file 1 (Section S1).

Delphi expert review (three rounds, n = 8). An eight-member interdisciplinary panel (clinical psychology, psychological counselling, psychometrics, social work, public health, community psychology, migration studies, and cross-cultural scale adaptation; 8 to 22 years of experience), none of whom had contributed to item generation, rated each item anonymously on essentiality, clarity, and representativeness [49]. The content validity ratio was computed as $CVR = (n_e - N/2)/(N/2)$ [50]; with $N = 8$ the retention threshold was $CVR \geq .75$, that is, seven of eight experts rating an item essential — the same critical value the exact binomial approach gives for panels of nine or fewer [51]. Item- and scale-level content validity indices followed Polit and Beck [52]. Items were eliminated for a CVR below threshold, redundancy, cross-loading or construct-contamination risk, risk of differential functioning by legal status or language proficiency, or measurement of an outcome rather than the latent trait. Round 1 reduced the 36-item pool to 21 (S-CVI/Average = .935); Round 2, an unscored discussion round, reduced it to 17; and Round 3, a dichotomous retain-or-remove vote, returned the final 16, of which ten had $CVR = 1.00$ and six $CVR = .75$ (S-CVI/Average = .953). The rounds also fixed the common instruction, which anchors responses to experiences following forced migration or displacement over approximately the past month, and the five response categories, 1 = *not at all true for me*, 2 = *not really true for me*, 3 = *somewhat true*, 4 = *quite true for me*, 5 = *completely true for me*. The resulting 16-item form was administered unchanged in the pilot and main studies (Additional file 1, Section S2).

Cognitive interviews (n = 12). Semi-structured interviews with 12 purposively sampled adults with forced migration experience (6 women; ages 20 to over 50; five from Syria, three Afghanistan, two Iraq, one Iran, one Ukraine; four reporting moderate, five good, and three very good Turkish comprehension) combined think-aloud with verbal probing at item and component level [53] and were coded under the four cognitive stages of comprehension, retrieval, judgment, and response formatting [54]. Two researchers independently coded all 192 item–participant units (Cohen's $\kappa = .87$). Items and structural components were evaluated against pre-specified elimination thresholds; none met a threshold, and the limited interpretation variation observed on five items fell within the intended content in every case. Turkish reading ability was established by researcher observation rather than by a standardized language test, which is a reproducibility limitation. The interview guide, including the item-level and structural probes, the coding framework and the a priori decision rules, is provided in Additional file 1 (Section S2).

Pilot study (N = 107) and sample size determination. The 16-item form was piloted with 107 forcibly displaced adults who did not take part in the main study; every questionnaire carried a unique participant code, and pilot codes were checked against the main-study records so that no pilot participant entered the main sample. The pilot matrix was factorable (Kaiser–Meyer–Olkin = .81; Bartlett's $\chi^2(120) = 755.33, p < .001$), parallel analysis supported three factors, internal consistency was adequate (ordinal $\alpha = .857$; $\omega = .905$; factor-level α

= .799–.866), and in a three-factor minimum-residual solution with oblimin rotation all 16 items loaded most strongly on their hypothesized factor (primary loadings .391–.843). Five items were flagged on descriptive pilot criteria, two of them already carrying an a priori monitoring flag recorded during the expert review; all five were carried forward and monitored across the main-study analyses. Target sample size was then set by Monte Carlo simulation from a deliberately misspecified population model parameterized on the pilot, each replication being fitted by a three-factor model estimated with diagonally weighted least squares using mean- and variance-adjusted test statistics (WLSMV) that omitted the misspecified terms. Across 500 replications per candidate size ($N = 250$ to 500 in steps of 50) every replication converged, and the proportion meeting $CFI > .95$, $TLI > .95$, $RMSEA < .06$ and $SRMR < .08$ simultaneously met or exceeded the 80% target at every candidate size but one (79.2% at $N = 400$), reaching 84.2% at the smallest size evaluated. The success rate was therefore not monotonic in N , a pattern attributable to Monte Carlo sampling variability across 500 replications rather than to any deterioration at larger N , since mean fit was comparable throughout. Because the target was already met at $N = 250$, that value is an upper bound on the required sample under this model rather than an estimate of it; the achieved sample substantially exceeded it (Additional file 1, Section S3).

Participants

The target population comprised adults (18+) who had left their country of origin because of forced migration and who could read and understand the language of administration, Turkish; the language criterion was established by self-report. Recruitment used snowball sampling in Istanbul, which hosts by a wide margin the largest population of Syrians under temporary protection, and Ankara, which concentrates the institutional and non-governmental service infrastructure through which displaced adults access support; because the measure includes a service and rights navigation dimension, sampling across settings differing in institutional density was substantively rather than logistically motivated. One feature of the Turkish context bears directly on that dimension: entitlements and the procedures attached to them differ by protection category, with Syrian nationals covered by temporary protection and Afghan and Iraqi nationals falling under the international protection regime, so the institutions and steps the third dimension refers to are not identical across origin groups.

The analytic sample comprised 583 adults, 57.8% men and 42.2% women, originating from Syria (49.7%), Afghanistan (27.7%), and Iraq (22.5%), and heterogeneous in age, education, length of stay, and self-reported Turkish comprehension. Full characteristics are given in Table 2, with percentages computed on valid responses (579 to 582 across variables). A subset of 141 participants (24.2%) provided retest data (Additional file 1, Section S5).

Table 2
Characteristics of the main-study analytic sample ($N = 583$)

Characteristic	<i>n</i>	%	Characteristic	<i>n</i>	%
Age group			Country or region of origin		
18–35	187	32.1	Syria	289	49.7
36–50	197	33.8	Afghanistan	161	27.7
51–65	122	21.0	Iraq	131	22.5
66 and above	76	13.1	Length of stay in Türkiye		
Gender			0–5 years	140	24.1
Woman	245	42.2	6–10 years	260	44.8
Man	335	57.8	11 years or more	180	31.0
Education level			Turkish comprehension		
None or primary	137	23.6	Moderate	158	27.3
Secondary or high school	224	38.6	Good	263	45.4
Bachelor's or graduate	220	37.9	Very good	158	27.3

Note. Percentages are calculated relative to the number of valid responses for each variable (valid $n = 579–582$); item-level missingness did not exceed 0.7% on any demographic variable, and category counts therefore do not sum to 583. The highest item-level rate on any of the 45 scale items was 0.9%, on rrs7. No participant selected the *other* gender category. Data were collected face-to-face in Istanbul and Ankara between 12 May and 12 June 2026. A further 141 participants from this sample completed a telephone retest approximately two weeks later.

Measures

RRS-16. Sixteen items (5-point; 1 = *not at all true for me*, 5 = *completely true for me*) with recall anchored to experiences following forced migration or displacement over approximately the past month: six items index individual adaptive capacity (F1), five support seeking and relational resource use (F2), and five service and rights navigation (F3). No item is reverse-scored and higher scores indicate greater mobilization capacity throughout. This includes rrs9 (*When I need support, I do not have difficulty asking for it*), the one negatively phrased item: agreement with it indicates greater rather than lesser capacity, so it is keyed in the same direction as the other fifteen and enters every score unrecoded. It was retained despite the double-negation risk it carries in translation, and its behaviour is monitored throughout the Results. Factor scores are the mean of the constituent items, computed when at least five of the six F1 items or four of the five F2 or F3 items are answered; a total score may be computed as a descriptive aggregate but should not carry the primary interpretation, for reasons given in the Results. The full item wording, factor assignments, and standardized loadings are reported with the confirmatory results. The full questionnaire as administered, including the participant information text, the common instruction, the response categories and the brief demographic form, is

provided in Additional file 1 (Section S2, Survey Draft), together with the administered Turkish wording and the English reference translation of each item (Appendix A, Table A1).

Brief Resilience Scale (BRS). ([13]; Turkish adaptation: [55]). Six items (5-point; items 2, 4 and 6 reverse-scored) assessing the capacity to recover from stress; administered as a convergent indicator. Ordinal $\alpha = .789$ [95% CI .756, .814].

General Self-Efficacy Scale (GSE). ([56]; Turkish: [57]). Ten items (4-point) assessing generalized confidence in coping with novel or difficult demands; administered as a convergent indicator, expected a priori to correlate positively with all three RRS-16 factors. Ordinal $\alpha = .819$ [.795, .838].

Four-item Perceived Stress Scale (PSS-4). ([58]; Turkish: [59]). Four items (5-point; items 2 and 3 reverse-scored) assessing the degree to which recent circumstances are appraised as unpredictable, uncontrollable, and overloading; administered as a discriminant indicator, since perceived stress is a distress appraisal rather than a resource. Ordinal $\alpha = .715$ [.677, .751], the lowest of the four but acceptable for a four-item form.

Patient Health Questionnaire depression module (PHQ-9). ([60]; Turkish: [61]). Nine items (4-point) screening depressive symptom severity over the preceding two weeks; administered as a discriminant indicator, since resilience must be distinguishable from the absence of symptoms. Ordinal $\alpha = .849$ [.829, .863].

Demographic and background form. Six variables: age group, gender, education level, country or region of origin, length of stay in Türkiye, and self-reported Turkish comprehension. Legal status was not recorded. In the pilot, questions touching on protection category visibly raised fear of deportation and reduced willingness to continue, so the item was dropped from the main-study form; this removes one grouping variable of substantive interest from the invariance analyses, and we return to the cost of that decision in the Discussion.

No social desirability measure was included.

Translation of the RRS-16 into English

The RRS-16 was developed and psychometrically validated in Turkish; the English version reported here was produced for international dissemination and was not itself administered in the present study. Translation followed the International Test Commission Guidelines for Translating and Adapting Tests [62] and COSMIN recommendations [47]. First, two independent bilingual translators, both native Turkish speakers with academic proficiency in English and familiarity with psychological measurement, produced independent forward translations. Second, the research team reconciled the two forward translations into a single synthesized version, resolving discrepancies through discussion and prioritizing conceptual and functional equivalence over literal correspondence. Third, one additional independent translator, a native English speaker blind to the source items, back-translated the synthesized version into Turkish. Fourth, an expert committee comprising the translators, a bilingual psychometrician, and a specialist in forced migration and mental health compared each back-translation against the Turkish source and evaluated semantic, idiomatic, experiential, and conceptual equivalence [63], revising items for which equivalence was not fully achieved; particular attention was given to rrs9, whose negated construction carries a double-negation risk in several of the languages into which the instrument is most likely to be adapted. Because all psychometric analyses reported here were conducted on the

Turkish administration, the English version should be regarded as a translated form whose cross-language measurement equivalence remains to be established.

Procedure

Data were collected face-to-face between 12 May and 12 June 2026 using standardized paper-and-pencil administration by trained field researchers, completed individually in quiet, semi-private settings (mean completion time 11.01 min, $SD = 2.04$). The five instruments were administered as intact blocks in a fixed order, but the items within each block were presented in a scrambled rather than a construct-grouped sequence, so that adjacent items did not belong systematically to the same factor or subscale. Two participants (0.3%) who reported adequate Turkish nonetheless encountered difficulty with the written form, reflecting the gap between spoken competence and reading fluency common where the language of settlement is acquired informally; for these two, interviewers read items aloud neutrally, confining clarification to procedural matters. Every other participant completed the form unaided, so administration mode was effectively uniform. Interviewers were trained to pause if visible distress emerged, to remind participants that participation was voluntary, and to direct participants to support resources where needed. No incentive was offered. Responses were confidential rather than anonymous: no names were recorded, each participant received a unique study code, and the master list linking codes to telephone numbers was stored separately. Participants who consented to re-contact were called approximately two weeks later and the measure was administered verbally by telephone, with responses matched by code; this mode difference is confounded with the retest interval and is treated as a limitation rather than a design feature. The study was approved on 26 December 2025 by the social and human sciences research ethics committee of the authors' institution, the approval covering all phases, each of which was conducted after that date; all procedures conformed to the Declaration of Helsinki, and participants provided written informed consent.

Four exclusion criteria were pre-specified and applied to the 592 raw records: failure of either of two embedded attention-check items; a missing-response rate of 10% or more across the 45 ordinal items administered in the session (RRS-16, 16; BRS, 6; GSE, 10; PSS-4, 4; PHQ-9, 9); straightlining; and a completion time that was missing or outside a plausible 5 to 90 minute range. Nine respondents (1.52%) met at least one criterion and were removed, yielding a final analytic sample of $N = 583$ (Additional file 1, Section S4).

Data analysis strategy

All RRS-16 indicators were treated as ordered-categorical [64], and no imputation was performed at any stage, with one documented exception: person-mean substitution was used solely to rank-order respondents for the upper–lower 27% discrimination computation, entering no other analysis and altering no reported item response (Additional file 1, Section S6).

The main sample was divided into calibration and validation subsamples by stratified random allocation on gender $\times$ age group; four participants with missing age or gender were excluded from the split only, leaving 579 for allocation (calibration $n = 290$, validation $n = 289$). Eleven equivalence tests across all demographic variables and all five scale totals were evaluated against a Bonferroni-corrected $\alpha = .05/12 = .00417$. The divisor was fixed at 12 a priori in the analysis script rather than set afterwards from the realized number of tests, which makes the correction marginally more conservative than an exact eleven-test correction (Additional file 1, Section S7).

Calibration analyses used polychoric exploratory factor analysis (minimum residual, oblimin), parallel analysis, Velicer's minimum average partial, the Kaiser criterion, and the scree elbow, together with exploratory graph analysis (graphical lasso with Walktrap community detection, Leiden as a sensitivity check, bootstrap $B = 200$) as a structurally independent check on dimensionality [65]. Graph-analytic estimation and every bootstrap procedure used complete cases ($n = 271$); the factor analysis used pairwise-present polychoric correlations. Two polychoric pipelines were run and both are archived; values reported below are from the primary pipeline, and the secondary pipeline recovered the same three-factor solution with fit indices differing by at most .004 and primary loadings by at most .028.

Validation analyses tested unidimensional, correlated three-factor, second-order, and bifactor models estimated with WLSMV under theta parameterization, with standardized latent variables, pairwise-present missing data, and robust scaled fit statistics throughout; target-rotated exploratory structural equation modelling served as a cross-loading sensitivity check. This four-model comparison, and the pairing of confirmatory factor analysis with an item-response framework on the same item set, follow the strategy now conventional in instrument validation [66]. Fit was evaluated against conventional cutoffs of CFI and TLI at or above .95, RMSEA at or below .06, and SRMR at or below .08 [67], with an RMSEA between .05 and .08 read as indicating reasonable fit [68] and the cutoffs themselves treated as guidelines rather than decision rules [69]. Under categorical estimation these indices tend to be more favourable than their maximum-likelihood counterparts, so a model falling short of the conventional values here would fall further short under maximum likelihood [70].

Full-sample analyses ($N = 583$) comprised (a) ordinal measurement invariance across six grouping variables, evaluated with directional change-in-fit criteria [71, 72] ($\Delta\text{CFI} \geq -.010$ and $\Delta\text{RMSEA} \leq .015$ at both steps, $\Delta\text{SRMR} \leq .030$ for metric versus configural and $\leq .010$ for scalar versus metric, minimum cell size 50 per group); (b) reliability (ordinal α with bootstrap 95% CI, ω total, composite reliability and average variance extracted, and bifactor indices); (c) test-retest stability (intraclass correlation, two-way random effects, absolute agreement); (d) convergent, discriminant, and known-groups validity in a joint latent model of the three RRS-16 factors and the four external measures; and (e) the secondary analyses of item response theory, network psychometrics, common method variance, differential item functioning, and estimator, item-removal, and residual-covariance sensitivity. Analyses were conducted in R (version 4.5.2) using lavaan (0.6–20), psych, EFAtools, EGAnet, semTools, BifactorIndicesCalculator, mirt, mokken, ordinal, bootnet, qgraph, networktools, mgm, irr, caret, and effectsize, with fixed seeds for all resampling (Additional file 1, Section S0).

Use of large language models. During manuscript preparation the authors used a large language model for language editing of the manuscript draft and for drafting portions of the analysis code. No large language model was used to generate hypotheses, conduct or interpret analyses, produce results, or select or summarize literature. All code was reviewed, executed, and verified by the authors, and all reported output was checked against the archived analysis files. The authors reviewed and verified all content and take full responsibility for the integrity of the work.

One correction is disclosed here. Hierarchical and specific omega coefficients were recomputed with a dedicated bifactor indices package [73] after an earlier version of the analysis script mis-specified the omega denominator, inflating hierarchical omega to .848 in the full sample and .836 in the validation subsample but affecting no other reported quantity. Corrected values are reported throughout.

Results

Phase 1 summary and item retention

Content validity was strong at both scored Delphi rounds. Round 1 reduced the 36-item pool to 21 items, all with item-level content validity indices of at least .875 (S-CVI/Average = .935); Round 2, an unscored discussion round, resolved the remaining redundancy and construct-overlap concerns and left 17 items; Round 3 returned a final 16 items, ten rated essential by all eight experts and six by seven of eight (S-CVI/Average = .953). Cognitive interviews confirmed comprehensibility: across 192 item–participant units two coders agreed at $\kappa = .87$, 14 units (7.3%) drew a flag from at least one coder, six were confirmed as genuine first-pass interpretation deviations, and no item and no structural component reached an elimination threshold. Eleven of the 16 items were rephrased in the intended direction by all 12 participants; the remaining five varied within the intended content and were retained unmodified.

In the pilot ($N = 107$), the matrix was factorable (Kaiser–Meyer–Olkin = .81; Bartlett's $\chi^2(120) = 755.33, p < .001$), the first four eigenvalues were 5.218, 2.468, 1.947 and 0.975, parallel analysis supported three factors, and internal consistency was adequate (ordinal $\alpha = .857$, $\omega = .905$; factor-level $\alpha = .799–.866$). All 16 items loaded most strongly on their hypothesized factor, with primary loadings from .391 to .843 and no cross-loading above .36.

In the calibration subsample ($n = 290$) the polychoric matrix was highly suitable for factor analysis (KMO = .899, item range .855–.951; Bartlett's $\chi^2(120) = 2045.59, p < .001$). Three of the four computable retention indices – parallel analysis, Velicer's minimum average partial, and the Kaiser criterion – indicated three factors; the automatic scree elbow indicated four, and the Hull method was not computable. With k fixed at three a priori, minimum-residual extraction with oblimin rotation fitted well (RMSEA = .034, 90% CI [.013, .051]; TLI = .978; RMSR = .025). Thirteen items loaded on their theoretical factor with primary loadings from .588 to .868. Three did not behave cleanly: rrs9 (.408) and rrs13 (.406) loaded on the theoretical factor but weakly, and rrs6 loaded at only .303 with a secondary loading of .203 and a communality of .303.

Exploratory graph analysis provided a structurally independent check. Graphical lasso estimation with Walktrap community detection recovered exactly three communities, and 15 of 16 items were assigned to the community matching their theoretical factor; the single exception was rrs6, placed with the F2 items. Leiden community detection recovered the same three communities and returned rrs6 to F1, so the misallocation is algorithm-specific rather than structural. Bootstrap analyses ($B = 200$) recovered $k = 3$ in 94.5% of replications, mean item-level stability was .930, and 12 of the 16 items were perfectly stable; the exceptions were rrs6 (.318), rrs9 (.753), rrs13 (.808) and rrs7 (.995). Bootstrap congruence for the exploratory solution averaged .988 ($SD = .003$), and sensitivity analyses across alternative extraction and rotation combinations reproduced the primary loading matrix exactly (Tucker's congruence = 1.000) (Additional file 1, Section S8).

Two facts structure everything that follows. Two dimensionality methods resting on different assumptions converged on the same three-factor solution; and the same three items – rrs6, rrs9 and rrs13 – were flagged at both the pilot and calibration stages. None reached an elimination threshold at either stage, so all 16 items were carried into the confirmatory stage and these three were tracked through it.

Confirmatory factor structure (validation subsample, $n = 289$)

Four competing models were estimated (Table 3). All model chi-square values were significant, as is expected at this sample size, so model comparison relied on descriptive fit indices and the information criteria. The one-factor model fitted poorly on every index, establishing that a single general dimension does not represent the item covariance structure. The correlated three-factor model met the conventional CFI criterion but not the conventional TLI or RMSEA criterion, and its SRMR met the conventional cutoff; its RMSEA of .064 falls within the range conventionally described as reasonable fit [68]. We therefore describe this model's fit as acceptable but not excellent, and make no claim that it meets the .95/.06 benchmarks. The second-order model was numerically identical to it: with exactly three first-order factors, three second-order loadings reproduce the three estimable factor correlations using the same number of parameters, so the two specifications are reparameterizations of one another rather than competing models. The bifactor model fitted best on every index at the cost of 13 additional parameters, and the two information criteria disagree about whether that price is worth paying – the Akaike information criterion (AIC) favours the bifactor model by 20.85, the Bayesian information criterion (BIC) favours the correlated model – which is what should happen when a more complex model buys real but modest improvement (Additional file 1, Section S9).

Table 3
Confirmatory factor analysis model fit comparison (validation subsample, $n = 289$).

Model	χ^2	df	CFI	TLI	RMSEA [90% CI]	SRMR	AIC	BIC
One-factor	842.84	104	.699	.653	.157 [.147, .167]	.131	12745.53	12921.52
Correlated three-factor	220.15	101	.952	.942	.064 [.052, .076]	.063	12209.61	12396.60
Second-order	220.15	101	.952	.942	.064 [.052, .076]	.063	12209.61	12396.60
Bifactor	136.53	88	.980	.973	.044 [.029, .058]	.044	12188.76	12423.41

Note. All models estimated with WLSMV and theta parameterization; χ^2 values are robust scaled statistics, all significant at $p < .001$. The second-order model is fit-equivalent to the correlated three-factor model in a three-factor application and is reported for completeness. AIC and BIC are undefined for WLSMV and are taken from the same four models refit by maximum likelihood with robust standard errors, purely as descriptive information criteria; they are not comparable with the WLSMV statistics in the same row. Conventional cutoffs: CFI and TLI $\geq .95$, RMSEA $\leq .06$, SRMR $\leq .08$; an RMSEA of .05 to .08 is conventionally read as reasonable fit [68]. No convergence failure or improper solution occurred in any model.

Standardized loadings on the retained correlated three-factor model ranged from .440 to .819, with rrs13 (.440) and rrs9 (.537) the only items below .58 (Fig. 1, Table 4). Latent factor correlations were moderate and well separated from unity ($F1-F2 = .500$, $F1-F3 = .429$, $F2-F3 = .484$). The second-order loadings were differentiated rather than uniform (.666, .751 and .644 for F1, F2 and F3), consistent with those correlations.

Note

Standardized WLSMV factor loadings on the indicator paths; factor intercorrelations shown as curved double-headed arrows. All loadings significant at $p < .001$. F1 = individual adaptive capacity; F2 = support seeking and

relational resource use; F3 = service and rights navigation.

Table 4
RRS-16 items, factor assignments, and standardized CFA factor loadings (validation subsample, $n = 289$).

Factor	Item	Item content	λ
Individual Adaptive Capacity (F1)	rrs1	When a new problem arises, I can find at least one practical step toward a solution.	.742
	rrs2	Even when things are uncertain, I am able to make decisions for myself and my family.	.790
	rrs3	When my plans fall through, I can make a new plan and adapt.	.706
	rrs4	Even when I feel overwhelmed, I can focus on what I am able to do and keep going.	.711
	rrs5	When I experience difficult emotions, I can find a way to calm myself.	.743
	rrs6	I believe I can improve my situation over time.	.589
Support Seeking and Relational Resource Use (F2)	rrs7	There is at least one person I can share my troubles with.	.753
	rrs8	In an emergency, there is at least one person I can rely on.	.789
	rrs9	When I need support, I do not have difficulty asking for it.	.537
	rrs10	I can build relationships with people from different groups (even if I do not fully speak their language).	.742
	rrs11	I can connect with my own community (compatriots/cultural circle).	.734
Service and Rights Navigation (F3)	rrs12	I can find a way to access the information I need (rights, institutions, processes).	.714
	rrs13	When I need healthcare, I know where to go.	.440
	rrs14	I know where to get support for official procedures (institutions, municipalities, NGOs, etc.).	.742
	rrs15	I can complete tasks such as appointments, applications, and forms on my own or with support.	.819
	rrs16	When I want to benefit from a right, I know what steps to take.	.750

Note. λ = standardized WLSMV loading; all $p < .001$. Items are administered in Turkish; the English wording is a translation for reporting. All items are scored in the same direction and none is reverse-scored. Items are numbered and grouped by factor for presentation; in administration they were interleaved so that consecutive items did not belong to the same factor.

Target-rotated exploratory structural equation modelling, which frees all cross-loadings, fitted excellently ($\chi^2(75) = 86.00$, $p = .181$; CFI = .996, TLI = .993, RMSEA = .023 [.000, .042], SRMR = .028) and every item retained its primary loading on its hypothesized factor. Two items showed a secondary loading at or above .20 and only one

exceeded .30. For rrs9 the primary (F2, .331) and secondary (F3, .327) loadings were effectively tied, and the ordering reversed under Geomin rotation (.325 versus .342), which indicates weak factorial assignment rather than a low but well-defined loading. Freeing the cross-loadings also pulled two primary loadings below conventional minimum adequacy: rrs13 fell to .277 and rrs6 to .319, against .440 and .589 in the confirmatory solution. This is the expected consequence of redistributing an item's variance across all three factors when its content is not cleanly factor-specific, but it is the strongest single piece of evidence against retaining these two items, and we return to it in the Discussion. Bootstrap re-estimation ($B = 200$) converged in every replication for both specifications.

Bifactor diagnostics settle how the instrument should be scored. In the full sample, common variance was divided almost evenly between a general factor and three comparatively balanced specific factors (explained common variance = .496), and hierarchical omega was .694, with specific omegas of .371, .376 and .412; the validation subsample gave the same picture (.464 and .669). Both indices fall below the benchmarks conventionally required to treat a multidimensional instrument as essentially unidimensional. The bifactor model is therefore retained as a diagnostic, not as the scoring model: the RRS-16 is scored as three correlated factors, and a total score is at most a descriptive aggregate.

Reliability

Internal consistency was good throughout (Table 5). Ordinal α was .884 for the 16-item total and .816 to .859 across the three factors, with ω total virtually identical in every case, and composite reliability (CR) between .829 and .868. Average variance extracted (AVE) exceeded .50 for F1 and F2; for F3 it was .498, marginally short of the conventional threshold, which we report rather than round up (Additional file 1, Section S12).

Test-retest stability over approximately two weeks ($n = 141$) was good for the total score, $ICC(2,1) = .779$ [.705, .836], and moderate for the three factors (.578 to .647). Two features of this estimate should temper its interpretation. The retest was administered verbally by telephone whereas baseline was paper-and-pencil, so mode is confounded with interval. And F3 showed a small but statistically significant mean decrease at retest ($-.131, p = .021$), indicating a modest systematic drift that matters for anyone using F3 to detect change; the total, F1 and F2 showed no significant mean shift. Minimal detectable change at 95% confidence, computed from these retest estimates, was 0.790 on the total score and 1.306 to 1.402 at factor level, which is the practical cost of the shorter factor scales. These are the values to use when interpreting change across two administrations; the smaller internal-consistency-based counterparts reported under Score distributions apply to a single administration.

Table 5
Reliability estimates for the RRS-16 and its three factors (full sample, $N = 583$; retest $n = 141$).

Scale	k	Ordinal α [95% CI]	ω total	CR	AVE	ICC(2,1) [95% CI]	SEM	MDC95
RRS-16 total	16	.884 [.868, .897]	.885	—	—	.779 [.705, .836]	.285	.790
Individual Adaptive Capacity (F1)	6	.859 [.843, .874]	.864	.868	.524	.647 [.540, .734]	.471	1.306
Support Seeking and Relational Resource Use (F2)	5	.824 [.800, .841]	.830	.834	.504	.578 [.457, .678]	.506	1.402
Service and Rights Navigation (F3)	5	.816 [.792, .838]	.824	.829	.498	.643 [.534, .731]	.477	1.322

Note. α and ω total computed from polychoric correlations (bootstrap $B = 200$). CR = composite reliability; AVE = average variance extracted, both from the three-factor CFA and therefore defined at factor level only. ICC(2,1) = two-way random effects, absolute agreement, single measures. SEM = standard error of measurement, computed as the pooled T1/T2 SD $\times \sqrt{(1 - \text{ICC})}$ in the retest subsample; MDC95 = minimal detectable change at 95% confidence ($1.96 \times \sqrt{2} \times \text{SEM}$), both on the 1–5 mean-score metric and both applicable to change across two administrations rather than to a single administration. Baseline administration was face-to-face and paper-and-pencil; retest was verbal by telephone.

Item-removal sensitivity. Because rrs9 and rrs13 were flagged repeatedly across stages, and because the Discussion considers whether either should be revised in future versions, we examined what their removal would do (Table 6). Removing rrs9 improves confirmatory fit measurably ($\Delta\text{CFI} = +.016$, $\Delta\text{RMSEA} = -.009$) and raises F2's model-based omega from .808 to .824. Removing rrs13 has almost no effect on fit ($\Delta\text{CFI} = +.004$, $\Delta\text{RMSEA} = +.001$) but raises F3's omega from .798 to .822. Removing both yields the best-fitting model of the four. Factor intercorrelations stayed within .40 to .50 in every scenario. We report this because it is the strongest available argument against our own retention decision, and we set out in the Discussion why we did not act on it. In brief: this analysis was itself run after observing rrs9's confirmatory performance, on the same subsample that produced it, so treating its result as grounds for removal would capitalize on chance rather than avoid it [74]. Both items are instead retained under continued monitoring, with rrs9 the higher priority (Additional file 1, Section S10).

Estimator and residual-covariance sensitivity. Two further pre-specified robustness checks are reported here. Refitting the three-factor model on the full sample by maximum likelihood with robust standard errors, which treats the responses as continuous, reproduced the categorical solution closely: the two sets of standardized loadings correlated at .993, the mean absolute difference was .037, thirteen of the sixteen items differed by less than .05, and both estimators met conventional thresholds (maximum likelihood: CFI = .972, RMSEA = .039; WLSMV: CFI = .962, RMSEA = .058; these full-sample values are not comparable with the validation-subsample figures in Table 3). Only rrs6 exceeded the .10 difference adopted for this check (.108), followed by rrs9 (.082) and rrs13 (.080), so the three items flagged throughout are also the three most estimator-dependent. A within-factor residual-correlation screen identified two pairs at or above the pre-specified $|r| = .10$ threshold, rrs9 with rrs11 ($-.114$) and rrs1 with rrs6 ($-.110$). Freeing both residual covariances improved fit only marginally ($\Delta\text{CFI} =$

.005, $\Delta\text{RMSEA} = -.003$) and left the loadings essentially unchanged (mean $|\Delta\lambda| = .008$, maximum .032), so the model is reported without them. Item-level discrimination, computed as the standardized difference between the upper and lower 27% of the total-score distribution, was large for every item (Cohen's $d = 1.27$ to 2.02) and corrected item-total correlations ranged from .387 to .581, with the lowest values in both cases on rrs13 and rrs9 (Additional file 1, Sections S12 and S17).

Table 6
Sensitivity of the confirmatory model to removal of rrs9 and rrs13 (validation subsample, $n = 289$).

Scenario	k	χ^2	df	CFI	TLI	RMSEA	SRMR	Factor ω affected
Full 16-item model (retained)	16	220.15	101	.952	.942	.064	.063	F2 = .808; F3 = .798
rrs9 removed	15	162.18	87	.968	.961	.055	.056	F2 = .824
rrs13 removed	15	192.61	87	.956	.946	.065	.061	F3 = .822
rrs9 and rrs13 removed	14	132.10	74	.974	.968	.052	.052	F2 = .824; F3 = .822

Note. All scenarios estimated with WLSMV and theta parameterization on the validation subsample using the same pipeline as the retained model. ω is model-based factor omega. F1's omega was .835 and unchanged in every scenario. In the final row the F2 and F3 omegas are those obtained in the corresponding single-removal models; factor omegas were not re-estimated for the 14-item model. Factor intercorrelations remained between .40 and .50 throughout. No item was removed from the published instrument; the analysis is reported as evidence bearing on the retention decision, which is discussed below.

Construct validity

Discriminant validity. All three heterotrait–monotrait ratios fell far below the .85 threshold, ranging from .518 to .546 in the full sample (listwise-complete $n = 547$; mean .533) and from .459 to .516 in the validation subsample. The Fornell–Larcker criterion was satisfied for every factor [75]: each $\sqrt{\text{AVE}}$ exceeded that factor's correlations with the other two (Table 7). The three factors are therefore empirically distinguishable, which is the claim the three-factor scoring rule depends on.

Convergent and divergent validity. Fully latent associations, estimated in a single seven-factor model that fitted well ($\chi^2(924) = 1059.91$, $p = .001$; CFI = .987, TLI = .986, RMSEA = .016, SRMR = .040), matched predictions in pattern and in relative magnitude (Table 7). All three factors correlated positively with both convergent measures and negatively with both divergent measures. F1 showed the strongest convergent associations, with the Brief Resilience Scale (.694) and the General Self-Efficacy Scale (.660), as expected of a factor indexing individual adaptive capacity. F3 related more strongly to general self-efficacy (.565) than to bounce-back resilience (.382), consistent with its focus on practical navigation. Eleven of the twelve associations were significant; the exception was F3 with perceived stress ($-.097$, $p = .064$). The divergent associations were uniformly weak, ranging from $-.097$ to $-.325$, which is the intended result: the RRS-16 is not a distress measure in disguise. One anomaly in the external measures should be noted because it bears on how the divergent evidence is read. Perceived stress and depressive symptoms correlated at only $-.019$ with each other, an implausibly low value for two constructs that ordinarily correlate positively and moderately. The anomaly is not confined to that pair: both distress measures also related weakly to both convergent measures (Brief Resilience Scale with PHQ-9 =

– .186, General Self-Efficacy Scale with PHQ-9 = – .180, Brief Resilience Scale with PSS-4 = – .168), whereas the convergent measures behaved as expected with each other and with the RRS-16. The attenuation therefore attaches to the distress instruments as a block rather than to any single pairing, and the divergent correlations reported here should be read as a lower bound of unknown tightness rather than as clean evidence of discriminability.

Table 7

Construct validity: fully latent external correlations, discriminant indices, and HTMT ratios (full sample, $N = 583$; HTMT on listwise-complete cases, $n = 547$).

Factor	r with BRS	r with GSE	r with PSS-4	r with PHQ-9	AVE	$\sqrt{\text{AVE}}$	Highest HTMT (pair)
Individual Adaptive Capacity (F1)	.694***	.660***	–.325***	–.286***	.524	.724	.546 (F2)
Support Seeking and Relational Resource Use (F2)	.541***	.460***	–.112*	–.167***	.504	.710	.546 (F1)
Service and Rights Navigation (F3)	.382***	.565***	–.097	–.147**	.498	.706	.535 (F1)

Note. External correlations are model-implied latent correlations from a joint seven-factor WLSMV model. BRS = Brief Resilience Scale; GSE = General Self-Efficacy Scale; PSS-4 = four-item Perceived Stress Scale; PHQ-9 = Patient Health Questionnaire depression module. AVE = average variance extracted, taken from the standalone three-factor model; in the joint seven-factor model the corresponding values are .538, .527 and .523. $\sqrt{\text{AVE}}$ is compared against that factor's correlations with the other two in the standalone model ($F1-F2 = .523$, $F1-F3 = .493$, $F2-F3 = .483$; joint-model values .522, .492 and .484), and exceeds them under either set. HTMT computed on listwise-complete cases from the full sample ($n = 547$); all values below the .85 and .90 thresholds. * $p < .05$, ** $p < .01$, *** $p < .001$.

Known-groups validity. Latent means were compared under scalar invariance across dichotomized forms of the six grouping variables: age 18–50 versus 51 and above, gender, education (none or primary versus secondary and above), origin (modal group versus other), and median splits on length of stay and Turkish comprehension. The differences were interpretable, and the single directional prediction was supported on one factor of three. The clearest effects were on service and rights navigation: participants with longer residence in Türkiye scored higher ($d = 0.422$, $p < .001$), as did the modal origin group ($d = 0.367$, $p < .001$) and those with higher self-reported Turkish proficiency ($d = 0.225$, $p = .011$). Longer residence also predicted higher support seeking and relational resource use ($d = 0.320$, $p < .001$). Older participants scored lower on individual adaptive capacity ($d = -0.274$, $p = .001$), as did participants with less education ($d = -0.244$, $p = .009$), and women scored slightly higher than men on support seeking ($d = 0.179$, $p = .020$). The one a priori directional hypothesis, that higher Turkish proficiency predicts higher scores, was therefore not confirmed as specified: it held for F3, was consistent in direction but not significant for F2, and was reversed and non-significant for F1 ($d = -0.110$, $p = .218$). We report it as the partial result it is. Collapsing a three-category proficiency self-report into a median split discards information and is one plausible reason the pattern is not uniform, so this comparison in particular should be repeated on the finer-grained variable. All other comparisons were exploratory. That length of stay and language proficiency move F3 while leaving F1 essentially unchanged is the pattern the three-factor structure predicts:

navigation capacity is acquired through exposure to a specific institutional system, whereas individual adaptive capacity is not (Additional file 1, Section S13).

Measurement invariance

The correlated three-factor model was tested for ordinal measurement invariance across six grouping variables using configural, metric and scalar steps. Five variables were estimable and all five supported full scalar invariance, with every change-in-fit index comfortably inside its criterion at both steps (Table 8); model fit was good throughout (CFI .959 to .975, RMSEA .042 to .060). Latent means are therefore comparable across gender, education, country or region of origin, length of stay, and Turkish comprehension.

Age group in its four-category form could not be evaluated. No participant in the 66-and-above subgroup selected the lowest response category on rrs14, and because threshold estimation under WLSMV requires every category to be observed within each group, that item's threshold structure was inestimable in that subgroup; the metric and scalar models are nested extensions and were therefore inestimable as well. Every other item showed complete category coverage in that subgroup, and rrs14 showed complete coverage in every other subgroup of every variable. This is a data-availability limitation of one item in one subgroup, not evidence of non-invariance. Under a dichotomized 18–50 versus 51-and-above contrast the empty cell disappears; the split was placed at 50 because it falls on an existing category boundary and yields the most balanced division available (384 versus 198). Configural, metric and scalar invariance were all supported under that contrast (Additional file 1, Section S11).

Table 8
Ordinal measurement invariance: nested change-in-fit comparisons (full sample, $N = 583$).

Grouping variable	Comparison	ΔCFI	$\Delta RMSEA$	$\Delta SRMR$	Decision
Gender	Metric – configural	+ .0021	– .0034	+ .0037	Supported
	Scalar – metric	– .0024	– .0017	– .0022	Supported
Education level	Metric – configural	+ .0063	– .0074	+ .0040	Supported
	Scalar – metric	– .0027	– .0018	– .0024	Supported
Country or region of origin	Metric – configural	+ .0038	– .0053	+ .0056	Supported
	Scalar – metric	– .0035	– .0016	– .0027	Supported
Length of stay in Türkiye	Metric – configural	+ .0026	– .0044	+ .0052	Supported
	Scalar – metric	+ .0011	– .0049	– .0036	Supported
Turkish comprehension	Metric – configural	+ .0073	– .0078	+ .0040	Supported
	Scalar – metric	– .0018	– .0032	– .0026	Supported
Age group (four categories)	Configural	–	–	–	Not estimable

Note. Metric step criteria: $\Delta CFI \geq -.010$, $\Delta RMSEA \leq +.015$, $\Delta SRMR \leq +.030$. Scalar step: $\Delta CFI \geq -.010$, $\Delta RMSEA \leq +.015$, $\Delta SRMR \leq +.010$. A positive ΔCFI indicates improved fit under the more constrained model. ΔTLI is not tabulated; at the scalar step it was positive for all five variables, diverging in sign from ΔCFI for gender, education, origin and Turkish comprehension, and in no case exceeded .006 in magnitude. Age group in its four-

category form was inestimable because one response category on rrs14 was unobserved in the 66-and-above subgroup; under a dichotomized age contrast all three steps were supported.

Secondary psychometric analyses

Item response theory and Mokken scaling. The generalized partial credit model outperformed the graded response model in all three blocks on both information criteria ($\Delta\text{AIC} = 6.9$ to 18.2). Discrimination parameters were all positive and ranged from 0.417 to 2.018 , with the three lowest values on the same items identified throughout: rrs13 (0.417), rrs6 (0.465) and rrs9 (0.488). Limited-information fit was acceptable in every block ($C2\ p = .063$ to $.917$; $\text{RMSEA} \leq 0.043$; $\text{SRMR} \leq 0.025$), and no item showed significant misfit after Bonferroni correction. Local independence requires care in interpretation. On the raw criterion, 16 of 35 within-factor item pairs exceeded $|Q3| \geq .20$ and all three blocks failed. But every Q3 residual was negative, ranging from -0.04 to -0.38 , which is the signature of the known downward bias of approximately $-1/(J-1)$ in short blocks – here -0.20 for F1 and -0.25 for F2 and F3. After bias correction only one pair survived, rrs8–rrs9 in F2. We report both figures because the raw count, taken alone, would misrepresent the instrument, and the corrected count, taken alone, would conceal that the correction was necessary. A second criterion is less favourable and is not subject to that bias: the likelihood-ratio LD χ^2 statistic exceeded 10 for 23 of the same 35 pairs (10 of 15 in F1, 9 of 10 in F2, 4 of 10 in F3) and ranks the three blocks differently from Q3. Some local dependence beyond the common factor therefore remains in the item set, and the bias-corrected Q3 count taken alone would understate it. Mokken analysis supported a single scale in each block with good scalability ($H = .448$ to $.478$; item $H_i .315$ to $.527$), automated item selection recovered exactly one scale per block, and no monotonicity or invariant-item-ordering violation occurred. Test information peaked between $\theta = -0.8$ and -0.1 , with conditional reliability at or above $.70$ spanning $\theta = -2.6$ to 1.9 for F1, -2.3 to 1.7 for F2 and -2.3 to 1.5 for F3 – that is, the instrument measures most precisely at and somewhat below the mean, which is where a screening instrument for a displaced population should be most precise (Additional file 1, Section S14).

Network psychometrics. A regularized partial correlation network estimated by graphical lasso ($\gamma = 0.5$, no thresholding, complete-case $n = 547$) retained 76 of 120 possible edges (sparsity = 0.367) and formed a single connected component. The topology mirrored the confirmatory structure closely: dense, strong within-factor edges and comparatively sparse, weaker between-factor edges (Fig. 2). Only two edges were negative and both were negligible ($|w| \leq .03$). Item rrs8 was the most central node on both strength and expected influence (1.130 on each) and rrs13 the least (strength 0.566 , expected influence 0.506); of the two centrality indices with adequate case-dropping stability, both reached $\text{CS} = 0.750$, whereas closeness (0.128) and betweenness (0.049) were unstable and are not interpreted. Bridge centrality identified rrs6 as by far the strongest connector across the three clusters (bridge strength = 0.417), followed by rrs9 (0.320) and rrs13 (0.291) – the same three items flagged by every other method, now with a substantive interpretation: they are weak indicators of their assigned factor precisely because their content spans factors. Clustering did not exceed that of comparable random graphs of the same density ($\text{Gamma} = 0.980$, $\text{Lambda} = 1.000$, $\sigma = 0.980$), so the topology is randomly clustered rather than small-world, no redundant item pair was detected, and the structure replicated across a split-half (edge-weight $r = 0.814$) and across the γ grid (edge-weight $r \geq 0.999$) (Additional file 1, Section S15).

Note

Node colour denotes the three CFA-derived factors, used here as a priori communities for bridge centrality. Edge thickness is proportional to the magnitude of the partial correlation. The display omits edges below $|w| = 0.03$, so

25 of the 76 estimated edges are not drawn; 22 of these are between-factor edges, so the contrast between dense within-cluster and sparse between-cluster connectivity is more pronounced in the figure than in the full estimated network. All reported edge counts and centrality statistics are computed from the complete network.

Common method variance. Four diagnostics were applied. Harman's single-factor test on the 45-item session-wide pool gave a first factor explaining 17.168% of variance, far below the 50% warning level. In the full sample the one-factor confirmatory model fitted poorly relative to the three-factor model (CFI .729 versus .962; $\Delta\text{CFI} = 0.233$; RMSEA lower by 0.094; the corresponding validation-subsample comparison in Table 3 gives $\Delta\text{CFI} = 0.253$), so item covariation is not adequately explained by a single common factor. Adding an orthogonal common latent factor to the three-factor model changed fit by $\Delta\text{CFI} = 0.026$, which exceeds the conventional 0.01 guideline, with a mean absolute loading change of 0.153 and an average shared variance of 18.9%. We report this as a failed diagnostic rather than a passed one. Two considerations bear on how much weight it should carry: in a single-instrument item set an orthogonal general factor absorbs substantive general resilience variance as well as method variance, and the three items whose loadings changed most (rrs13, rrs6, rrs9) are the same three whose factorial assignment is weakest by every other method, which is what one would expect if the common factor were absorbing genuine cross-domain content. Heterotrait–monotrait ratios among the three factors remained between .518 and .546 ($n = 547$), so the inter-factor relations are not dominated by a single method-driven factor (Additional file 1, Section S16).

Differential item functioning. Ordinal logistic regression screened all 96 item-by-group combinations (16 items $\times$ 6 grouping variables), each grouping variable dichotomized as in the known-groups analyses, with matching on the mean of the remaining items in the same factor and no group level falling below the minimum of 30 per level required for a test. No combination met the pre-specified criterion for practically meaningful DIF ($\Delta\text{McFadden } R^2 \geq .010$). One combination reached statistical detectability – uniform DIF on rrs1 across the dichotomized age contrast, Holm-adjusted $p = .019$ – with a magnitude of $\Delta R^2 = .0064$, well below the practical threshold. Because age enters here in its dichotomized form, this result is not a re-test of the four-category age grouping that was inestimable in the invariance analysis. Maximum uniform effects ranged from .0018 to .0064 across the six variables and maximum non-uniform effects from .0017 to .0044. Item-level comparability across the examined groups is therefore supported without item-specific corrections (Additional file 1, Section S17).

Score distributions

Full-sample means on the 1–5 metric were closely similar across the three factors: F1 $M = 3.239$ ($SD = 0.787$), F2 $M = 3.201$ ($SD = 0.780$; $n = 582$, one participant falling below the completeness rule for that factor), F3 $M = 3.260$ ($SD = 0.781$), with a total-scale mean of 3.234 ($SD = 0.617$), reported as a descriptive aggregate only, for the reasons given under the bifactor diagnostics. Distributions were centred and well spread, with total scores running from 1.562 to 4.875 and interquartile ranges of roughly one scale point at factor level. No item approached the 20% floor or ceiling criterion (maxima 10.4% and 16.2%), and no factor distribution approached the scale bounds, which matters for an instrument intended to detect variation within a population often assumed to be uniformly low in resource access. Single-administration measurement error on the mean metric, computed from internal consistency rather than from the retest and using complete-case alphas that differ from those in Table 5 by at most .003, was 0.207 for the total score and 0.295 to 0.334 at factor level, giving minimal detectable change values of 0.574 and 0.817 to 0.926 respectively. These assume temporal stability and are accordingly smaller than the retest-based values in Table 5, which remain the ones to use for change across two administrations. They are usable for group-level monitoring, but too coarse for individual change decisions on a

single factor, which is the honest limit of a 16-item instrument (Additional file 1, Sections S12 and S17, and Appendix A).

Discussion

The RRS-16 was developed to measure what forcibly displaced adults can mobilize – within themselves, through their relationships, and through the institutions that govern their status – rather than how far they bounce back from adversity in general. Across a three-round expert review, cognitive interviews with displaced adults, a pilot, and a main study of 583 Syrian, Afghan and Iraqi adults in Türkiye, the instrument recovered a replicable three-factor structure, achieved good internal consistency, and supported scalar invariance across five grouping variables. Four findings warrant interpretation.

Before those, the four pre-specified hypotheses can be adjudicated. H1 (structure) was supported: exploratory factor analysis and exploratory graph analysis independently recovered three factors in the calibration subsample, the correlated three-factor model fitted acceptably in the validation subsample and the one-factor model did not, and although the bifactor model fitted better, its general factor was too weak to justify a total score. H2 (reliability) was supported for internal consistency and for the stability of the total score, but only partially at factor level, where two-week intraclass correlations were moderate rather than good. H3 (invariance) was supported for the five grouping variables that could be estimated, and no item showed practically meaningful differential functioning; the four-category age contrast was inestimable, so for that variable the hypothesis is untested rather than disconfirmed. H4 (validity) was supported in its ordering – individual adaptive capacity strongest throughout, service and rights navigation weakest on distress and more strongly tied to self-efficacy than to dispositional resilience – but its divergent half rests on evidence we report as inconclusive, and one of the twelve associations did not reach significance. The single directional known-groups prediction, on Turkish proficiency, held for service and rights navigation only.

What the three dimensions measure, and how they differ from what already exists

The first dimension, individual adaptive capacity, is the part of the construct that existing instruments already capture. Its items index problem-focused action under uncertainty, plan flexibility and emotional self-regulation, and they behaved as that reading predicts: this factor correlated most strongly of the three with the Brief Resilience Scale and with general self-efficacy, and it was the only factor unmoved by how long a participant had lived in Türkiye. Domain-general resilience measures such as the Connor–Davidson scale [11] and the Brief Resilience Scale [13] measure this territory competently, and nothing in our data suggests otherwise. The case for a new instrument does not rest here.

The second dimension, support seeking and relational resource use, is the one most easily mistaken for perceived social support, and the distinction matters. Its items ask not whether support exists but whether the respondent can reach it – whether there is someone to rely on in an emergency, whether asking is difficult, whether relationships can be built across a language barrier and maintained within one's own community. This is the mobilization side of Hobfoll's conservation-of-resources account [17] rather than the availability side, and it deliberately spans Putnam's bonding and bridging distinction [43] in a population for whom the two are not

interchangeable: co-national ties supply continuity, cross-group ties supply access. The dimension moved with length of residence in Türkiye, as an acquired relational repertoire should.

The third dimension, service and rights navigation, is the one no existing resilience instrument for displaced adults models as a dimension of its own, and it is the analytic centre of this paper. Its items concern knowing where to go for healthcare, how to complete an application, and what steps an entitlement requires. Instruments developed within displacement contexts approach this territory from other directions without occupying it: the Child and Youth Resilience Measure locates resilience in the caregiving ecology of adolescents rather than in adult institutional competence [28]; the Self-Reliance Index measures the household economic outcome of successful navigation rather than the capacity that produces it [32]; and the Refugee Post-Migration Stress Scale measures the bureaucratic obstacle itself rather than the respondent's ability to work through it [33]. The closest approach comes from outside the resilience literature altogether: a healthcare navigation competency scale developed for refugees in the United States covers the same kind of capability at far greater length, but within one service system, as a standalone instrument rather than a component of a resource-mobilization construct, and with its published evaluation limited to content and face validity [34]. Our claim is therefore narrower than it may first appear, and we state it in its narrow form: what the RRS-16 adds is not the discovery that navigation can be measured, but its integration as a third, group-comparable dimension alongside individual and relational capacity in a single short instrument.

The distinction is not merely conceptual. Service and rights navigation related considerably more strongly to general self-efficacy than to bounce-back resilience, which is what one expects of a practical capability belief rather than a recovery trait. That pattern also invites an alternative reading, which should be stated rather than avoided: if this factor correlates .565 with general self-efficacy and only .382 with dispositional resilience, perhaps it is domain-specific self-efficacy under another name. Three considerations argue against collapsing it. General self-efficacy items index confidence across unspecified demands, whereas these items name a specific institutional object — an application, an entitlement, a route to care. The factor remained discriminable from general self-efficacy in the joint latent model rather than approaching redundancy with it. And it moves with length of residence and language proficiency, which is what an acquired institutional competence should do and what a generalized efficacy belief has no particular reason to do. We nonetheless accept that the boundary between a domain-specific efficacy belief and a navigational capacity is thin, and that administering a purpose-built institutional self-efficacy measure alongside the RRS-16 would test it directly. The dimension is grounded in the post-migration stressor literature [5] and in Ager and Strang's account of integration as access to functioning systems [44] rather than in the trauma-recovery tradition.

Why the three dimensions cannot be collapsed

If the three factors correlate moderately, why not report a single resilience score? Bifactor diagnostics answer that question directly. Explained common variance was close to even between a general factor and three balanced specific factors, and hierarchical omega fell well short of the level at which a multidimensional instrument may be treated as essentially unidimensional. Substantial reliable variance therefore sits in the specific factors, and a total score would discard it. The network analysis pointed the same way from a different direction: the item system formed three dense clusters corresponding to the confirmatory factors, joined by comparatively sparse between-cluster edges. We therefore score the RRS-16 as three correlated factors and treat the total as a descriptive aggregate only.

The known-groups results show what that decision buys. Length of residence in Türkiye moved service and rights navigation by the largest margin observed in these comparisons ($d = 0.422$) and support seeking by $d = 0.320$, and higher Turkish proficiency moved navigation by a smaller but reliable margin ($d = 0.225$), while individual adaptive capacity was left essentially untouched – the standardized difference by length of stay was 0.002. These comparisons rest on dichotomized groupings in a single cross-sectional sample, so what they establish is the asymmetry rather than the magnitude of any one coefficient. An instrument confined to the F1 territory, which is what a domain-general resilience measure is, would therefore be blind to precisely the variation that changes as displaced adults settle. That asymmetry is the empirical case for the RRS-16: not that general resilience measures are invalid, but that in this population they cannot see the part of the construct that responds to time, language and institutional exposure, and that is the part practitioners and policy makers can act on.

Why no item was removed

Three items – rrs6, rrs9 and rrs13 – were flagged repeatedly, and our sensitivity analysis shows that removing rrs9 would improve confirmatory fit and raise the reliability of its factor. We nevertheless retained all sixteen items, and the reasoning should be set out in full.

First, under the model the instrument is actually scored with, none of the three items met the criteria that would identify a psychometrically inadequate item. Every standardized loading in the confirmatory solution, including rrs13's, remained above conventional minimum-adequacy thresholds [76]; no residual correlation reached the magnitude ordinarily taken to indicate serious local misfit, rrs9's largest approaching but not reaching it; and each item retained a positive, interpretable discrimination parameter under item response modelling. Two qualifications belong in the same paragraph rather than in a footnote, because both cut against us. Under target-rotated exploratory structural equation modelling – not the scoring model, but the more permissive one – the primary loadings of rrs13 (.277) and rrs6 (.319) fall at or below those same thresholds. And the local-dependence evidence is criterion-dependent: bias-corrected Q3 leaves a single dependent pair, whereas the likelihood-ratio criterion flags 23 of the 35 within-factor pairs. Neither result reaches a pre-specified removal threshold, and neither is what inadequacy conventionally means, but together they place the distance between weak and adequate at less than the confirmatory solution alone would suggest. Weak is not the same as inadequate, and the evidence still places these three items in the former category – with rrs6 and rrs13 nearer the boundary than their confirmatory loadings imply.

Second, removing items on the strength of loadings observed in a single sample is the classic route to a structure that does not replicate [74]. The concern is sharpened, not resolved, by our own sensitivity analysis: it was run after observing rrs9's confirmatory performance, on the same subsample that produced it, so acting on its result would repeat the capitalization-on-chance problem in one further pass over the same data rather than escape it. What the analysis establishes is the size of the trade-off, not a warrant for revision; an independent sample is required for that.

Third, removal would trade psychometric tidiness against content validity that was independently established. Both rrs6 and rrs9 were judged by the expert panel to carry conceptually distinct facets – improvement expectancy for rrs6, the absence of a barrier to asking for help for rrs9 – and neither drew a comprehension problem in the cognitive interviews, where rrs9 was among the five items on which interpretation varied within the intended content and rrs6 was rephrased in the intended direction by all twelve participants; current best-

practice guidance holds that expert-panel and target-population evidence should inform, not be overridden by, subsequent psychometric performance [77, 78].

Fourth, and most instructive, the three flagged items were the three strongest bridge nodes in the item network. They connect the clusters. On that reading they load weakly on their assigned factor not because they are poor items but because their content genuinely spans domains — believing one's situation can improve is not purely individual, and asking for help is not purely relational. Removing them would raise fit indices by removing the very content that makes the three dimensions parts of one construct.

What follows is a monitoring obligation rather than a revision. rrs9 is the priority: its primary and secondary loadings were effectively tied under exploratory structural equation modelling and the ordering reversed under an alternative rotation, which indicates weak factorial assignment rather than a low but well-defined loading — a problem that reassignment or rewording could address as readily as deletion, and one that should be settled in an independent sample rather than in this one. rrs6 and rrs13 follow, for a related reason: their primary loadings drop below conventional adequacy once cross-loadings are freed, which is what content spanning domains looks like in a rotated solution and which rewording could also address. All three should be re-examined before any revision of the instrument is published (Additional file 1, Section S18).

Practical use

Because scalar invariance held across gender, education, origin, length of stay and Turkish comprehension, latent means may be compared across those groups, which is the property that makes the instrument usable for needs assessment rather than description alone. The three-factor profile is the unit of interpretation: two people with the same total score may differ entirely in whether their constraint is internal, relational or institutional, and the intervention implied by each differs. Service providers can use the F3 profile to identify clients whose difficulty is navigational rather than psychological — a distinction that determines whether the appropriate response is counselling or case management. The measurement error we report is adequate for group-level monitoring and for cohort comparison; it is too coarse for decisions about an individual on a single factor, and we do not recommend that use (Additional file 1, Appendix A).

Limitations

Nine limitations bound these conclusions. First, all data come from one country, one language and two provinces, recruited by snowball sampling, a procedure that over-samples people embedded in reachable networks and is therefore liable to inflate the relational and navigational dimensions in particular — precisely the dimensions on which the instrument is meant to detect variation; the English wording reported here was produced for dissemination and has not been administered, so cross-language equivalence is entirely unestablished and the instrument should not be used in another language without full adaptation and revalidation. Second, legal status was not recorded. For a measure whose third dimension concerns rights navigation this is the most consequential omission in the design, since protection category plausibly conditions both access and the perception of access, and we cannot test it. The omission was not an oversight: pilot participants met the question with evident fear of deportation, and we judged that retaining it would compromise recruitment and the candour of the remaining answers. That judgement protected the sample at the cost of the single most informative grouping variable, and it leaves open whether the third dimension indexes navigational capacity or the protection category that conditions it. Marital status, employment and household economic level

were likewise not recorded, and each is a plausible confound for the relational and navigational dimensions. Third, the test–retest estimate is confounded: the retest was administered verbally by telephone whereas baseline was self-completed on paper, so mode and interval cannot be separated, and the small but significant decline on service and rights navigation at retest may reflect either genuine drift or that difference. The interval also sits inside the instrument’s recall window: respondents are asked about approximately the past month, whereas the two administrations were about two weeks apart, so the occasions refer to partly overlapping periods and the intraclass correlations should be read as an upper bound on stability. Fourth, the common latent factor diagnostic exceeded the conventional guideline. We report this as a failed test. It is not decisive – an orthogonal general factor in a single-instrument item set absorbs substantive general capacity as well as method variance, and the items whose loadings shifted most were the three whose content is genuinely cross-domain – but readers should weigh it, and no social desirability measure was administered to adjudicate. Relatedly, all items are positively keyed, which leaves an acquiescence risk that the diagnostics we ran are not designed to detect; future versions should include a reverse-keyed subset. Fifth, the divergent evidence rests on two instruments that behaved oddly, and not only with one another: perceived stress and depressive symptoms correlated at close to zero, and both related weakly to the two convergent measures as well. Whatever produced this – translation, response style, or the suitability of these instruments in this population – it attenuates every correlation involving the distress block, so the divergent correlations can bear less weight than their pattern suggests, and replication with distress measures validated in displaced populations is required. Sixth, invariance across the four-category age variable could not be estimated, because one response category on a single item was unobserved in the oldest subgroup; under a dichotomized contrast all steps were supported, but the finer comparison remains untested. Seventh, the validation subsample yields roughly three and a half observations per free parameter, below the textbook recommendation, and the confirmatory results should be read with that in mind even though every model converged admissibly. Eighth, the known-groups comparisons and the differential item functioning screen were run on dichotomized groupings, including a median split of a three-category proficiency self-report; this discards information, is the most likely reason the one directional hypothesis was supported on a single factor, and means the reported effect sizes should be read as coarse. Ninth, eligibility required self-reported ability to read and understand Turkish. This restricts range on the third dimension by design, since every item in that dimension concerns acting within Turkish-language institutions and the adults least able to do so were not sampled. The finding that higher proficiency raises navigation scores is therefore established over a truncated range and is to some degree built into the sampling frame; it should be read as consistent with what the dimension is meant to capture rather than as an independent test of it.

Beyond these, criterion validity and responsiveness were not assessed: no gold standard exists for this construct, and no longitudinal or intervention data were collected, so the instrument’s sensitivity to genuine change is unknown. One form of that gap is worth naming, because it is the one a reader is most likely to raise. The obvious criterion for a resilience measure is symptom level, and incremental prediction of distress over an existing resilience scale would be the obvious test of added value. That test is the wrong one here: the dimension this instrument adds indexes the capacity to act within institutions rather than the distress that follows from being unable to, which is why H4 predicted – and the data show – that service and rights navigation relates least strongly of the three factors to distress. The criteria that would discriminate this dimension are behavioural and administrative: service contact, unmet need, help-seeking, and completion of procedures. None was collected. What stands in their place is the known-groups asymmetry, in which length of

residence and language proficiency move navigation while leaving individual adaptive capacity unchanged – the pattern an acquired institutional competence should show and a distress proxy should not.

Conclusions

The RRS-16 measures resource mobilization capacity in forcibly displaced adults across three dimensions – individual adaptive capacity, support seeking and relational resource use, and service and rights navigation – with good internal consistency, a replicable factor structure, and scalar invariance across five demographic groupings. Its distinctive contribution is the third dimension, which general resilience instruments do not capture and which varies with settlement time and language competence. The scale is suitable for group-level assessment and comparison in Turkish-speaking displaced populations; use in other languages requires cross-cultural adaptation and revalidation.

Abbreviations

AIC	
Akaike information criterion	
AVE	
average variance extracted	
BIC	
Bayesian information criterion	
BRS	
Brief Resilience Scale	
CFI	
comparative fit index	
CI	
confidence interval	
COSMIN	
COnsensus-based Standards for the selection of health Measurement INstruments	
CR	
composite reliability	
CVR	
content validity ratio	
GSE	
General Self-Efficacy Scale	
ICC	
intraclass correlation coefficient	
PHQ-9	
Patient Health Questionnaire depression module	
PSS-4	
four-item Perceived Stress Scale	
RMSEA	
root mean square error of approximation	
RRS-16	

Refugee Resilience Scale
SEM
standard error of measurement
SRMR
standardized root mean square residual
STROBE
Strengthening the Reporting of Observational Studies in Epidemiology
TLI
Tucker–Lewis index
WLSMV
diagonally weighted least squares with mean- and variance-adjusted test statistics.

Declarations

Ethics approval and consent to participate

All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments. The study protocol was approved by the Ethics Committee for Social and Human Sciences of the Scientific Research and Publication Ethics Board of Gümüşhane University, Türkiye (meeting dated 26 December 2025, decision no. 2025/10; reference no. E-76551937-108.99-395340). The approval covered all phases of the study, including item generation, the expert panel review, cognitive interviewing, the pilot study, and the main study. Written informed consent was obtained from all individual participants included in the study, including consent to be re-contacted for the retest administration. Participants were informed that participation was voluntary, that they could skip any question or discontinue at any time without penalty, and how their data would be stored, protected, and shared.

Consent for publication

Not applicable. The manuscript contains no data, images, videos, or personal details relating to an identifiable individual.

Availability of data and materials

The anonymized datasets generated and analysed during the current study, together with all analysis scripts, output tables, codebooks, and study instruments, are available in the Open Science Framework repository at <https://osf.io/jv8mz/> (anonymized view-only link for peer review: https://osf.io/jv8mz/overview?view_only=635464c7dff343c2ad99a3e8cf7da7aa). The RRS-16 may be used and adapted for non-commercial research and service-evaluation purposes without the authors' written permission, provided that this article is cited; the Turkish administration form and the English reporting translation are deposited in the same repository. The questionnaire as administered, the cognitive interview guide and the full supplementary material are also provided as Additional file 1.

Competing interests

The authors declare that they have no competing interests.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Authors' contributions

Öznur Çınar: conceptualization; methodology; investigation; formal analysis; writing – original draft; writing – review and editing. Ekrem Cengiz: conceptualization; methodology; formal analysis; software; data curation; writing – review and editing. Emel Yıldız: methodology; investigation; validation; writing – review and editing. Salih Yıldız: conceptualization; supervision; data curation; resources; writing – review and editing. Büşra Tosunoğlu Kusan: investigation; data curation; visualization; writing – review and editing. All authors read and approved the final version of the manuscript and agree to be accountable for all aspects of the work.

Acknowledgements

The authors thank the eight members of the expert panel who took part in the three-round modified Delphi review; the displaced adults who gave their time to the cognitive interviews, the pilot study, and the main study; and the field interviewers.

Authors' information

Öznur Çınar, PhD, is a lecturer in the Department of Emergency Aid and Disaster Management at Bayburt University, Türkiye. She has almost nine years of experience in the field of disaster management. Her research interests include disaster, hazard and risk management, disaster preparedness, and disaster and emergency prevention and response.

Ekrem Cengiz, PhD, is a professor in the Department of Business Administration at Gümüşhane University, Türkiye. He has almost eight years of experience in the field of disaster management and twenty years of experience in business administration. His research interests include disaster, hazard and risk management, disaster resilience, disaster and emergency prevention and response, marketing, statistics, management and organization, and leadership.

Emel Yıldız, PhD, is a professor in the Department of Business Administration at Gümüşhane University, Türkiye. She has fifteen years of experience in business administration. Her research interests include disaster preparedness, disaster and emergency prevention and response, service marketing, consumer behaviour, and word-of-mouth marketing.

Salih Yıldız, PhD, is a professor in the Department of Business Administration at Gümüşhane University, Türkiye. He has twenty years of experience in business administration. His research interests include brand management, risk management, marketing management, and marketing strategies.

Büşra Tosunoğlu Kusan, PhD, is a professor in the Department of Business Administration at Gümüşhane University, Türkiye. She has sixteen years of experience in business administration. Her research interests include accounting, management accounting, and sustainability.

Use of large language models

During manuscript preparation, the authors used a large language model for language editing of the manuscript draft and for drafting portions of the analysis code. No large language model was used to generate hypotheses, conduct or interpret analyses, produce results, or select or summarize literature. All code was reviewed, executed, and verified by the authors, and all reported output was checked against the archived analysis files. The authors reviewed and verified all content and take full responsibility for the integrity of the work.

Preprint. The manuscript has not been posted as a preprint, has not been published previously, and is not under consideration by any other journal.

References

1. World Health Organization. World report on the health of refugees and migrants. Geneva: World Health Organization. 2022. <https://www.who.int/publications/i/item/9789240054462>. Accessed 14 Aug 2026.
2. Steel Z, Silove D, Phan T, Bauman A. Long-term effect of psychological trauma on the mental health of Vietnamese refugees resettled in Australia: a population-based study. *Lancet*. 2002;360:1056–62. 10.1016/S0140-6736(02)11142-1.
3. Nguyen TP, Uribe Guajardo MG, Sahle BW, Renzaho AMN, Slewa-Younan S. Prevalence of common mental disorders in adult Syrian refugees resettled in high income Western countries: a systematic review and meta-analysis. *BMC Psychiatry*. 2022;22:15. 10.1186/s12888-021-03664-7.
4. Silove D. The psychosocial effects of torture, mass human rights violations, and refugee trauma: toward an integrated conceptual framework. *J Nerv Ment Dis*. 1999;187:200–7.
5. Miller KE, Rasmussen A. War exposure, daily stressors, and mental health in conflict and post-conflict settings: bridging the divide between trauma-focused and psychosocial frameworks. *Soc Sci Med*. 2010;70:7–16. 10.1016/j.socscimed.2009.09.029.
6. Nowak AC, Nutsch N, Brake TM, Gehrlein LM, Razum O. Associations between postmigration living situation and symptoms of common mental disorders in adult refugees in Europe: updating systematic review from 2015 onwards. *BMC Public Health*. 2023;23:1289. 10.1186/s12889-023-15931-1.
7. Grupp F, Mewes R, Tudorache A-C. Trauma, post-migration stress, and mental health among forcibly displaced people in Germany: does resilience have a protective effect? *Int J Soc Psychiatry*. 2026;72:311–24. 10.1177/00207640251358094.
8. Santambrogio J, Cimminiello N, Wisidagamage Don P, Leon E, Miragliotta E, Capuzzi E, et al. Influence of post-migration living difficulties on mental health among refugees and asylum seekers: a scoping review on clinical tools. *Int J Soc Psychiatry*. 2024;70:1191–201. 10.1177/00207640241251748.
9. Rutter M. Psychosocial resilience and protective mechanisms. *Am J Orthopsychiatry*. 1987;57:316–31. 10.1111/j.1939-0025.1987.tb03541.x.
10. Werner EE, Smith RS. *Vulnerable but invincible: a longitudinal study of resilient children and youth*. New York: McGraw-Hill; 1982.
11. Connor KM, Davidson JRT. Development of a new resilience scale: the Connor-Davidson Resilience Scale (CD-RISC). *Depress Anxiety*. 2003;18:76–82. 10.1002/da.10113.
12. Campbell-Sills L, Stein MB. Psychometric analysis and refinement of the Connor-Davidson Resilience Scale (CD-RISC): validation of a 10-item measure of resilience. *J Trauma Stress*. 2007;20:1019–28. 10.1002/jts.20271.

13. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The Brief Resilience Scale: assessing the ability to bounce back. *Int J Behav Med.* 2008;15:194–200. 10.1080/10705500802222972.
14. Wagnild GM, Young HM. Development and psychometric evaluation of the Resilience Scale. *J Nurs Meas.* 1993;1:165–78.
15. Friborg O, Hjemdal O, Rosenvinge JH, Martinussen M. A new rating scale for adult resilience: what are the central protective resources behind healthy adjustment? *Int J Methods Psychiatr Res.* 2003;12:65–76. 10.1002/mpr.143.
16. Ungar M. The social ecology of resilience: addressing contextual and cultural ambiguity of a nascent construct. *Am J Orthopsychiatry.* 2011;81:1–17. 10.1111/j.1939-0025.2010.01067.x.
17. Hobfoll SE. Conservation of resources: a new attempt at conceptualizing stress. *Am Psychol.* 1989;44:513–24. 10.1037/0003-066X.44.3.513.
18. Hobfoll SE. The influence of culture, community, and the nested-self in the stress process: advancing conservation of resources theory. *Appl Psychol.* 2001;50:337–421. 10.1111/1464-0597.00062.
19. Ewunie TM, Hayen A, Dawson A. Measurement tools used to assess individual health assets among refugee populations: a scoping review. *Health Promot Int.* 2025;40:daae199. 10.1093/heapro/daae199.
20. Windle G, Bennett KM, Noyes J. A methodological review of resilience measurement scales. *Health Qual Life Outcomes.* 2011;9:8. 10.1186/1477-7525-9-8.
21. Velickovic K, Rahm Hallberg I, Axelsson U, Borrebaeck CAK, Ryden L, Johnsson P, et al. Psychometric properties of the Connor-Davidson Resilience Scale (CD-RISC) in a non-clinical population in Sweden. *Health Qual Life Outcomes.* 2020;18:132. 10.1186/s12955-020-01383-3.
22. Pulido-Martos M, Fernandez-Sanchez MD, Lopez-Zafra E. Measurement invariance across gender and age in the Connor-Davidson Resilience Scale (CD-RISC) in a Spanish general population. *Qual Life Res.* 2020;29:1373–84. 10.1007/s11136-019-02389-1.
23. Kovalenko J, Mahgoub E, Nasr RA. Resilience and coping behavior among war refugees from diverse ethnic backgrounds. *Discov Psychol.* 2026;6:137. 10.1007/s44202-026-00698-4.
24. Lane R, Taylor H, Ellis F, Rushworth I, Chiu K. Resilience and its association with mental health among forcibly displaced populations: a systematic review and meta-analyses. *J Affect Disord.* 2025;379:387–400. 10.1016/j.jad.2025.03.015.
25. Fadhli TN, Sauter DA, Doosje B. Adversity, emotion, and resilience among Syrian refugees in the Netherlands. *BMC Psychol.* 2022;10:257. 10.1186/s40359-022-00963-w.
26. Demirci H, Gecer E, Goktas SS, Balaban G. Social cognition and connectedness in adult refugees: the case of Syrians living in Turkiye. *BMC Psychol.* 2025;13:1032. 10.1186/s40359-025-03361-0.
27. Sundvall M, Titelman D, DeMarinis V, Borisova L, Cetrez O. Safe but isolated: an interview study with Iraqi refugees in Sweden about social networks, social support, and mental health. *Int J Soc Psychiatry.* 2021;67:351–9. 10.1177/0020764020954257.
28. Panter-Brick C, Hadfield K, Dajani R, Eggerman M, Ager A, Ungar M. Resilience in context: a brief and culturally grounded measure for Syrian refugee and Jordanian host-community adolescents. *Child Dev.* 2018;89:1803–20. 10.1111/cdev.12868.
29. Clark JN, Jefferies P. Measuring resilience and the importance of resource connectivities: revising the Adult Resilience Measure (RRC-ARM). *Soc Sci.* 2023;12:290. 10.3390/socsci12050290.

30. Terrana A, Bruno W, Ibrahim N, Kaiser BN, Wei J, Al-Delaimy W. The Somali Distress and Resilience Scale: development of a novel measure for Somali adults. *Transcult Psychiatry*. 2024;61:915–27. 10.1177/13634615241272982.
31. Getnet B, Alem A. Construct validity and factor structure of sense of coherence (SoC-13) scale as a measure of resilience in Eritrean refugees living in Ethiopia. *Confl Health*. 2019;13:3. 10.1186/s13031-019-0185-1.
32. Seff I, Leeson K, Stark L. Measuring self-reliance among refugee and internally displaced households: the development of an index in humanitarian settings. *Confl Health*. 2021;15:56. 10.1186/s13031-021-00389-y.
33. Malm A, Tinghog P, Narusyte J, Saboonchi F. The Refugee Post-Migration Stress Scale (RPMS): development and validation among refugees from Syria recently resettled in Sweden. *Confl Health*. 2020;14:2. 10.1186/s13031-019-0246-5.
34. Yeo S, Lee I, Ehiri J, Magrath P, Ernst K, Kim YR, et al. Developing and validating a HEalthCare NAavigation Competency (HECNAC) Scale for refugees in the United States. *PLoS ONE*. 2025;20:e0314057. 10.1371/journal.pone.0314057.
35. Aroian KJ, Schappler-Morris N, Neary S, Spitzer A, Tran TV. Psychometric evaluation of the Russian language version of the Resilience Scale. *J Nurs Meas*. 1997;5:151–64. 10.1891/1061-3749.5.2.151.
36. Antonovsky A. *Unraveling the mystery of health: how people manage stress and stay well*. San Francisco: Jossey-Bass; 1987.
37. Luthar SS, Cicchetti D, Becker B. The construct of resilience: a critical evaluation and guidelines for future work. *Child Dev*. 2000;71:543–62. 10.1111/1467-8624.00164.
38. Southwick SM, Bonanno GA, Masten AS, Panter-Brick C, Yehuda R. Resilience definitions, theory, and challenges: interdisciplinary perspectives. *Eur J Psychotraumatol*. 2014;5:25338. 10.3402/ejpt.v5.25338.
39. Bandura A. Guide for constructing self-efficacy scales. In: Pajares F, Urdan T, editors. *Self-efficacy beliefs of adolescents*. Greenwich: Information Age Publishing; 2006. pp. 307–37.
40. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. *Behav Res Ther*. 2006;44:1–25. 10.1016/j.brat.2005.06.006.
41. Snyder CR. Hope theory: rainbows in the mind. *Psychol Inq*. 2002;13:249–75. 10.1207/S15327965PLI1304_01.
42. Cohen S, Wills TA. Stress, social support, and the buffering hypothesis. *Psychol Bull*. 1985;98:310–57. 10.1037/0033-2909.98.2.310.
43. Putnam RD. *Bowling alone: the collapse and revival of American community*. New York: Simon and Schuster; 2000.
44. Ager A, Strang A. Understanding integration: a conceptual framework. *J Refug Stud*. 2008;21:166–91. 10.1093/jrs/fen016.
45. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Lancet*. 2007;370:1453–7. 10.1016/S0140-6736(07)61602-X.
46. American Educational Research Association, American Psychological Association. *National Council on Measurement in Education. Standards for educational and psychological testing*. Washington, DC: American Educational Research Association; 2014.

47. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol*. 2010;10:22. 10.1186/1471-2288-10-22.
48. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quinonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front Public Health*. 2018;6:149. 10.3389/fpubh.2018.00149.
49. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs*. 2000;32:1008–15.
50. Lawshe CH. A quantitative approach to content validity. *Pers Psychol*. 1975;28:563–75. 10.1111/j.1744-6570.1975.tb01393.x.
51. Ayre C, Scally AJ. Critical values for Lawshe's content validity ratio: revisiting the original methods of calculation. *Meas Eval Couns Dev*. 2014;47:79–86. 10.1177/0748175613513808.
52. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. *Res Nurs Health*. 2006;29:489–97. 10.1002/nur.20147.
53. Willis GB. *Cognitive interviewing: a tool for improving questionnaire design*. Thousand Oaks: Sage; 2005.
54. Tourangeau R, Rips LJ, Rasinski K. *The psychology of survey response*. Cambridge: Cambridge University Press; 2000.
55. Dogan T. Adaptation of the Brief Resilience Scale into Turkish: a validity and reliability study. *J Happiness Well-Being*. 2015;3:93–102.
56. Schwarzer R, Jerusalem M. Generalized Self-Efficacy Scale. In: Weinman J, Wright S, Johnston M, editors. *Measures in health psychology: a user's portfolio. Causal and control beliefs*. Windsor: NFER-NELSON; 1995. pp. 35–7.
57. Aypay A. The adaptation study of the General Self-Efficacy Scale into Turkish. *Inonu Univ J Fac Educ*. 2010;11:113–32.
58. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav*. 1983;24:385–96. 10.2307/2136404.
59. Eskin M, Harlak H, Demirkiran F, Dereboy C. The adaptation of the Perceived Stress Scale into Turkish: a reliability and validity analysis. *New Symp J*. 2013;51:132–40.
60. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16:606–13. 10.1046/j.1525-1497.2001.016009606.x.
61. Sari YE, Kokoglu B, Balcioglu H, Bilge U, Colak E, Unluoglu I. Turkish reliability of the Patient Health Questionnaire-9. *Biomed Res*. 2016;27(Spec Iss 2):460–2.
62. International Test Commission. *The ITC guidelines for translating and adapting tests*. 2nd ed. Hemel Hempstead: International Test Commission. 2017. https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf. Accessed 14 Aug 2026.
63. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000;25:3186–91. 10.1097/00007632-200012150-00014.
64. Rhemtulla M, Brosseau-Liard PE, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol Methods*. 2012;17:354–73. 10.1037/a0029315.

65. Golino HF, Epskamp S. Exploratory graph analysis: a new approach for estimating the number of dimensions in psychological research. *PLoS ONE*. 2017;12:e0174035. 10.1371/journal.pone.0174035.
66. Jovanovic V, Lazic M. Vaccination Attitudes Examination (VAX) Scale: a bifactor-ESEM approach in a youth sample (15–24 years). *BMC Psychol*. 2023;11:351. 10.1186/s40359-023-01382-1.
67. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Model*. 1999;6:1–55. 10.1080/10705519909540118.
68. Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. *Testing structural equation models*. Newbury Park: Sage; 1993. pp. 136–62.
69. Marsh HW, Hau KT, Wen Z. In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Struct Equ Model*. 2004;11:320–41. 10.1207/s15328007sem1103_2.
70. Xia Y, Yang Y, RMSEA, CFI, TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. *Behav Res Methods*. 2019;51:409–28. 10.3758/s13428-018-1055-2.
71. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. *Struct Equ Model*. 2007;14:464–504. 10.1080/10705510701301834.
72. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct Equ Model*. 2002;9:233–55. 10.1207/S15328007SEM0902_5.
73. Dueber DM. BifactorIndicesCalculator: bifactor indices calculator. R package version 0.2.2. 2017. 10.13023/edp.tool.01
74. MacCallum RC, Roznowski M, Necowitz LB. Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychol Bull*. 1992;111:490–504. 10.1037/0033-2909.111.3.490.
75. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res*. 1981;18:39–50. 10.1177/002224378101800104.
76. Tabachnick BG, Fidell LS. *Using multivariate statistics*. 7th ed. Boston: Pearson; 2019.
77. Stefana A, Damiani S, Granzio U, Provenzani U, Solmi M, Youngstrom EA, et al. Psychological, psychiatric, and behavioral sciences measurement scales: best practice guidelines for their development and validation. *Front Psychol*. 2025;15:1494261. 10.3389/fpsyg.2024.1494261.
78. Morgado FFR, Meireles JFF, Neves CM, Amaral ACS, Ferreira MEC. Scale development: ten main limitations and recommendations to improve future research practices. *Psicol Reflex Crit*. 2017;30:3. 10.1186/s41155-016-0057-1.

Figures

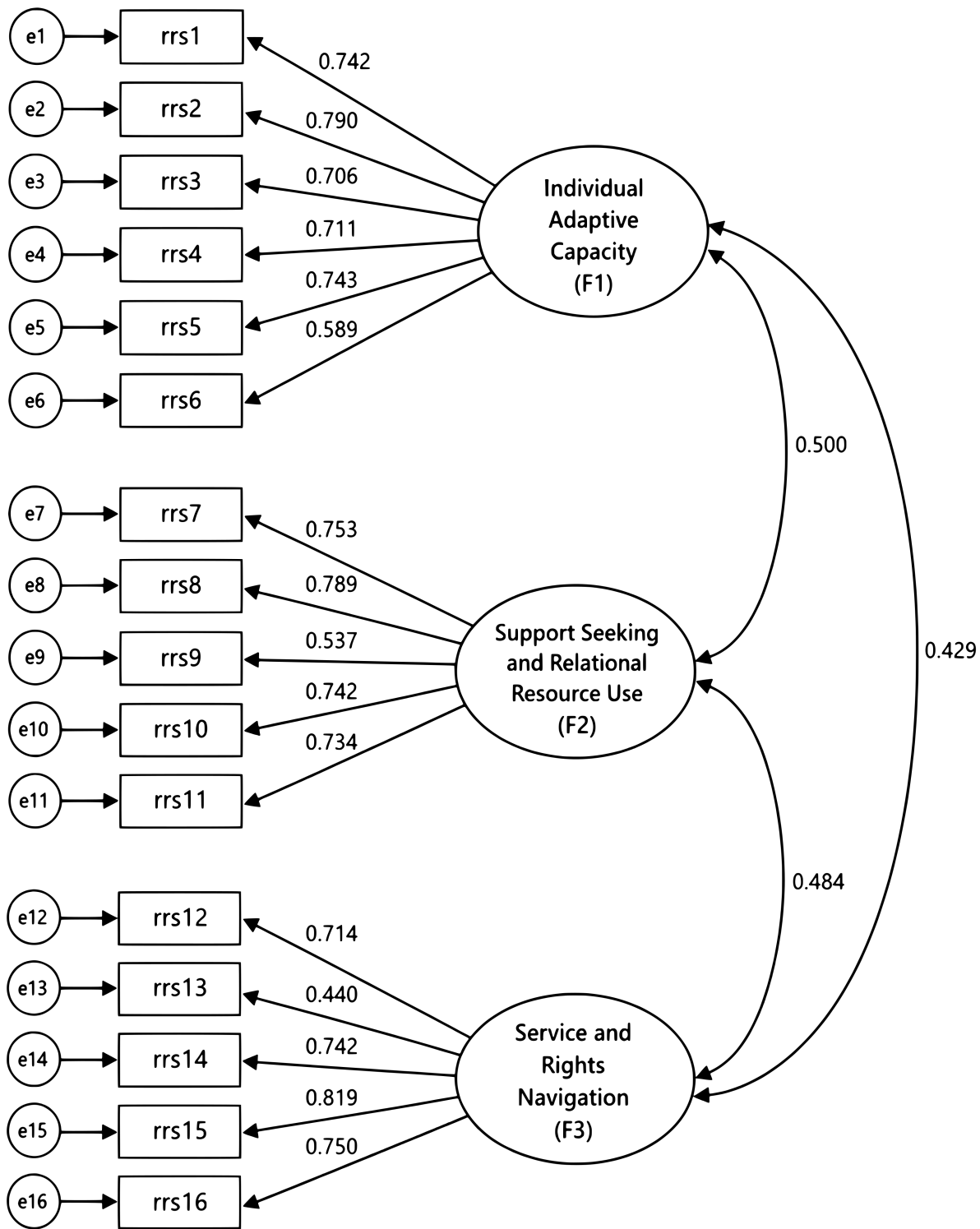


Figure 1

Path diagram of the correlated three-factor CFA model (RRS-16; validation subsample, $n= 289$).

Note. Standardized WLSMV factor loadings on the indicator paths; factor intercorrelations shown as curved double-headed arrows. All loadings significant at $p < .001$. F1 = individual adaptive capacity; F2 = support seeking and relational resource use; F3 = service and rights navigation.

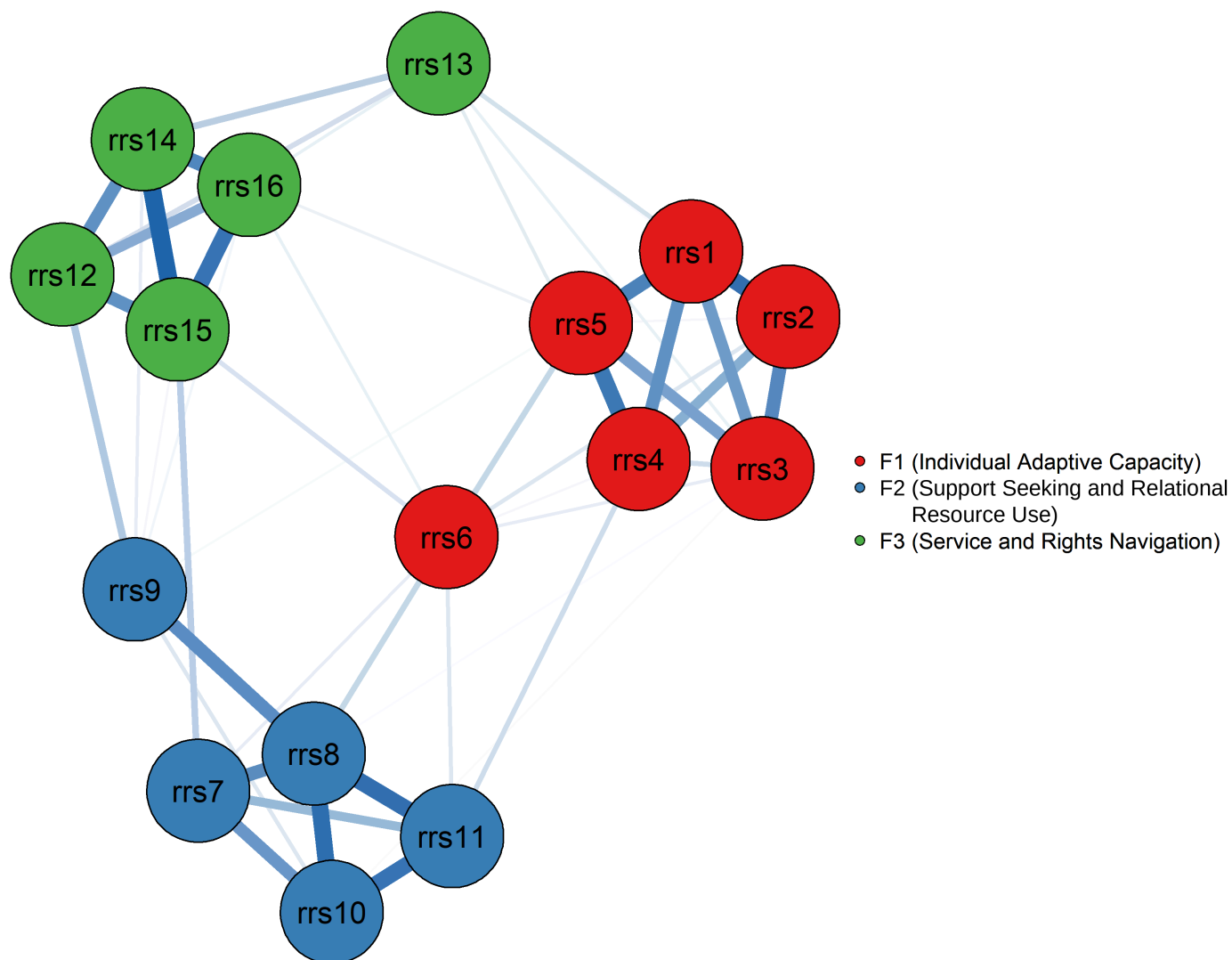


Figure 2

Regularized partial correlation network of the 16 RRS-16 items (graphical lasso, $\gamma = 0.50$, no thresholding; complete-case $n = 547$).

Note. Node colour denotes the three CFA-derived factors, used here as a priori communities for bridge centrality. Edge thickness is proportional to the magnitude of the partial correlation. The display omits edges below $|w| = 0.03$, so 25 of the 76 estimated edges are not drawn; 22 of these are between-factor edges, so the contrast between dense within-cluster and sparse between-cluster connectivity is more pronounced in the figure than in the full estimated network. All reported edge counts and centrality statistics are computed from the complete network.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)