Scale Development & Validation in R

From item screening to factor structure, reliability and scoring

Rverse Analytics

A reliable total score is not automatically a valid measurement model. Define the construct, response process and intended use before item reduction; use separate evidence for dimensionality, reliability and validity.

Development map

Stage	Primary question	Evidence / R tool
Construct definition	What is and is not measured?	theory, interviews, content map
Item generation	Does content cover every domain?	expert review, cognitive interviewing
Pilot audit	Do options and items behave as intended?	frequencies, missingness, floor/ceiling
Dimensionality	How many latent dimensions?	parallel analysis, EFA, CFA
Reliability	How precise are scores?	omega, alpha, test-retest, conditional SE
Validity	Do scores relate as predicted?	convergent, discriminant, criterion evidence
Invariance	Is meaning comparable across groups/time?	multigroup CFA / DIF

Item audit and reverse scoring

```
library(dplyr)
library(psych)

items <- d |> select(starts_with("item_"))

# Keep the original variables; create explicit scored copies
reverse_1_to_5 <- function(x) 6 - x
scored <- items |>
  mutate(
    item_03 = reverse_1_to_5(item_03),
    item_08 = reverse_1_to_5(item_08)
  )

audit <- tibble(
  item = names(scored),
  n = sapply(scored, function(x) sum(!is.na(x))),
  missing_pct = 100 * sapply(scored, function(x) mean(is.na(x))),
  mean = sapply(scored, function(x) mean(x, na.rm = TRUE)),
  sd = sapply(scored, function(x) sd(x, na.rm = TRUE)),
  floor_pct = 100 * sapply(scored, function(x) mean(x == 1, na.rm = TRUE)),
  ceiling_pct = 100 * sapply(scored, function(x) mean(x == 5, na.rm = TRUE))
)
```

Confirm response bounds, reverse-key logic, straight-lining, duplicated items and whether missingness clusters by item position or subgroup.

Correlations and factorability

```
# Ordinal Likert items: polychoric correlation is usually preferable
poly <- psych::polychoric(scored)$rho

KMO(poly)
cortest.bartlett(poly, n = nrow(scored))

# Empirical guide to factor count; do not use eigenvalue > 1 alone
fa.parallel(poly, n.obs = nrow(scored),
            fa = "fa", fm = "minres", plot = TRUE)
```

Check	What it tells you	Caution
KMO	shared variance adequate for factoring	not a validity test
Bartlett	correlation matrix differs from identity	often significant in large samples
Parallel analysis	observed eigenvalues vs random data	combine with theory and interpretability
Residual matrix	local areas of poor fit	inspect item wording/dependence

Exploratory factor analysis

```
efa <- fa(
  r = poly,
  nfactors = 3,
  n.obs = nrow(scored),
  fm = "minres",
  rotate = "oblimin"
)
print(efa$loadings, cutoff = 0.30, sort = TRUE)
efa$Phi          # factor correlations
efa$communality
efa$residual
```

Use an oblique rotation when constructs may correlate. Review primary loadings, cross-loadings, communalities, residuals and content coverage together; do not delete items mechanically from one cutoff.

Confirmatory factor analysis

```
library(lavaan)

model <- '
  Physical =~ item_01 + item_02 + item_03 + item_04
  Emotional =~ item_05 + item_06 + item_07 + item_08
  Social =~ item_09 + item_10 + item_11 + item_12
'

fit <- cfa(
  model,
  data = scored,
  ordered = names(scored),
  estimator = "WLSMV",
  std.lv = TRUE,
  missing = "pairwise"
)

summary(fit, standardized = TRUE, fit.measures = TRUE)
fitMeasures(fit, c("chisq", "df", "cfi", "tli", "rmsea", "srmr"))
standardizedSolution(fit)
modindices(fit, sort. = TRUE, maximum.number = 10)
```

Modification indices generate hypotheses, not permission to chase fit. Add correlated residuals or cross-loadings only with content justification and validate them in new data.

Reliability and score construction

```
# Alpha includes item diagnostics; omega reflects a factor model
alpha_out <- psych::alpha(scored, check.keys = FALSE)
omega_out <- psych::omega(poly, nfactors = 3, plot = FALSE)

alpha_out$total[c("raw_alpha", "std.alpha", "average_r")]
alpha_out$item.stats[, c("r.drop", "mean", "sd")]

# Score only after the retained structure and missing-item rule are fixed
d$scale_mean <- rowMeans(scored, na.rm = TRUE)
d$n_answered <- rowSums(!is.na(scored))
d$scale_mean[d$n_answered < 10] <- NA_real_
```

Report omega/alpha with confidence intervals where possible. Very high reliability may signal redundant items; “alpha if deleted” is not a standalone deletion rule.

Cross-validation, validity and invariance

- Split or independently collect EFA and CFA samples; avoid confirming a model in the same data used to discover it.
- Prespecify convergent and discriminant correlations with direction and plausible magnitude.
- For known-groups or criterion validity, report effect estimates and CIs, not significance alone.
- Test configural, metric and scalar invariance before comparing latent means across groups/time.
- Examine differential item functioning when ordinal items may behave differently across subgroups.

Reporting checklist

- Describe construct definition, item source, expert/respondent review and scoring direction.
- Report sample, missingness, item distributions, correlation type, extraction and rotation.
- Give factor-count evidence, loadings, cross-loadings, communalities and model-fit indices.

- Separate dimensionality, reliability and validity claims; provide uncertainty.
- Publish the final items, scoring rule, missing-item rule and version identifier.
