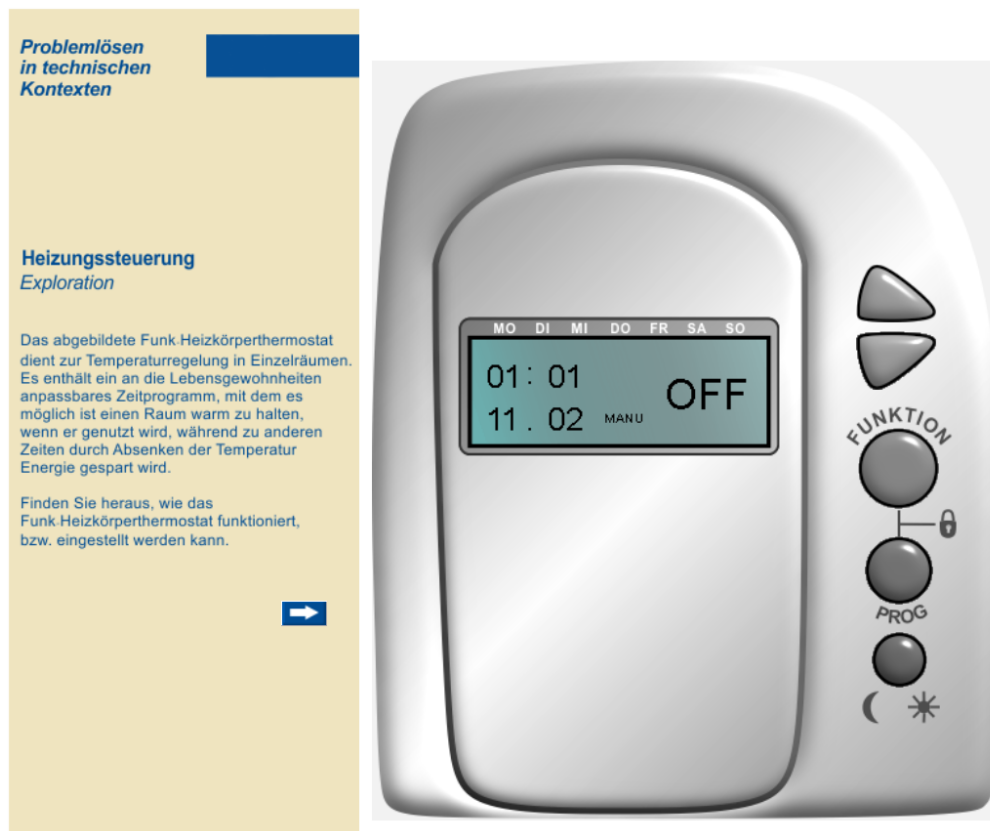


Elektronisches Supplement 2. TPL-Test und Skalierung

Das elektronische Supplement 2 beschreibt eine Beispielaufgabe des simulationsbasierten Instruments von Stemmann (2016), mit dem Fähigkeiten beim technischen Problemlösen (TPL) erfasst werden können, sowie die Modellierung von TPL als Personenfähigkeit anhand von Methoden der Item-Response-Theorie (IRT).

Test des Technischen Problemlösens (TPL). Die ESM Abbildung 2.1 zeigt die Aufgabe „Heizungssteuerung“, in der Personen mit dem abgebildeten Gerät per Mausclick interagieren und verschiedene Handlungen ausführen können (z. B. Datum und Uhrzeit einstellen, Raumtemperatur erhöhen, Zeiten programmieren). Im Test werden zwei Phasen unterschieden: Zunächst kann ein Gerät ausführlich ausprobiert werden (10-minütige Explorationsphase; Instruktion: „Finden Sie heraus, wie das Gerät funktioniert bzw. eingestellt werden kann.“). Anschließend soll es nach einer vorgegebenen Problemstellung in einen bestimmten Systemzustand gebracht werden (5-minütige Steuerungsphase; Instruktion: „Stellen Sie das Heizungsthermostat so ein, dass von Montags bis Freitags ab 07.00 Uhr auf eine Temperatur von 19°C geheizt wird.“).



ESM Abbildung 2.1. Steuerungsaufgabe „Heizungssteuerung“ aus dem Test für die Erfassung des technischen Problemlösens.

Elektronisches Supplement 2. TPL-Test und Skalierung

Vorgehen zur Skalierung. Für die Skalierung wurden zunächst die relativen Lösungshäufigkeiten sowie punktbiseriale Korrelationen als Maß der Trennschärfe für alle Items betrachtet. Anschließend wurde zur Testwertermittlung ein Rasch-Modell auf die dichotomen Itemantworten angepasst (Embretson & Reise, 2013). Dabei wurden *expected a posteriori* (EAP) Schätzer als Testwertschätzer verwendet. Zur Beurteilung der Passung des Rasch-Modells wurden der Item-Infit und Item-Outfit (Schwellenwerte: 0.7-1.3; Wright & Linacre, 1994) sowie Q3-Statistiken (Schwellenwert: $|\text{Wert}| > .2$; Chen & Thissen, 1997) untersucht. Für jedes Item wurden zudem die beobachteten Item-Response Kurven gemeinsam mit den modellimplizierten itemcharakteristischen Kurven auf Abweichungen und unerwartete Asymptoten visuell inspiziert. Die Analysen wurden in R 4.1.0 (R core Team, 2021) mit *TAM* (Robitzsch, Kiefer & Wu, 2021) und *irtos* (Partchev, Maris & Hattori, 2017) umgesetzt.

Ergebnisse. Aufgrund einer zu geringen Trennschärfe ($r_{pb} = 0.00$) wurde Item 13 (Mobiltelefon) von der weiteren Analyse ausgeschlossen (siehe ESM Tabelle 2.1). Die EAP-Reliabilität konnte dadurch von .52 auf .56 verbessert werden. Item 4 (Receiver) wies zwar ebenfalls eine geringe Trennschärfe auf ($r_{pb} = 0.06$). Auf einen Ausschluss wurde dennoch verzichtet, da die Konstruktrepräsentativität nicht weiter eingeschränkt werden sollte. Zudem war eine deutliche Verbesserung der EAP-Reliabilität nicht gegeben (ebenfalls .56).

ESM Tabelle 2.1. Deskriptive Itemstatistiken sowie Item-Infit und Outfit.

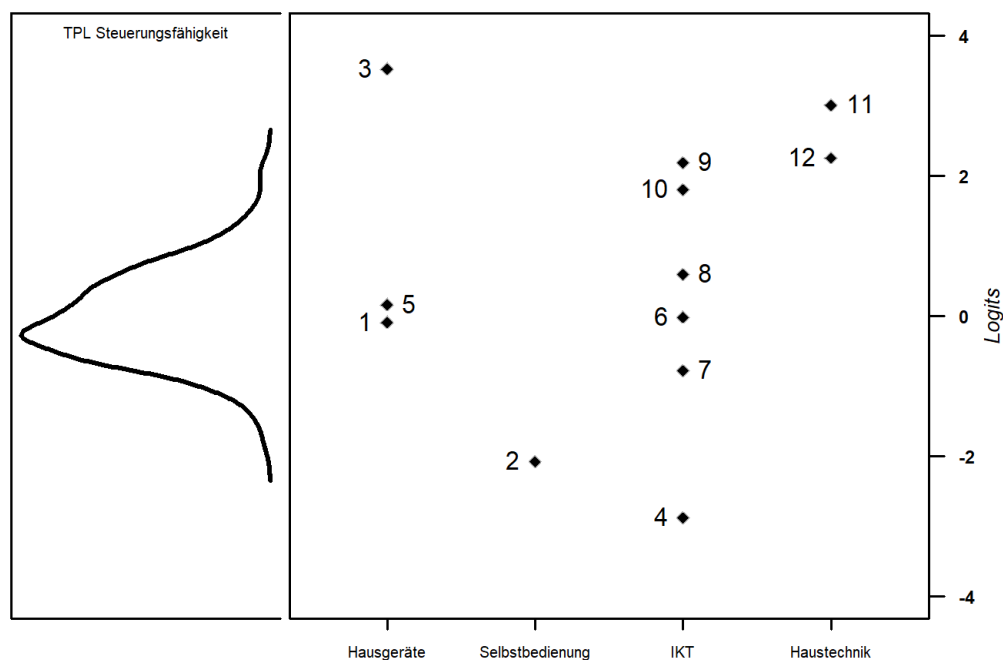
Item	Simulation	Bereich	% korrekt	r_{pb}	Infit	Outfit
1	Waschmaschine	Hausgerätetechnik	52	.18	1.05	1.05
2	Parkscheinautomat	Selbstbedienungsautomat	86	.20	1.00	0.98
3	Spülmaschine	Hausgerätetechnik	4	.15	1.01	0.92
4	Receiver	IKT	93	.06	1.04	1.14
5	Kaffeevollautomat	Hausgerätetechnik	47	.23	1.03	1.03
6	Küchenradio	IKT	50	.24	1.02	1.02
7	Navigationsgerät	IKT	66	.23	1.02	1.01
8	Fotodrucker	IKT	38	.22	1.03	1.04
9	Radiowecker	IKT	13	.34	0.93	0.79
10	Festnetztelefon	IKT	17	.32	0.96	0.89
11	Rollladensteuerung	Haustechnik	6	.16	1.00	1.01
12	Heizungssteuerung	Haustechnik	12	.40	0.89	0.72

Anmerkungen: r_{pb} ist die punktbiseriale Korrelation des Items mit dem Summenwert exkl. des Items. IKT = Informations- und Kommunikationstechnologie

Elektronisches Supplement 2. TPL-Test und Skalierung

Die ESM Abbildung 2.2 stellt die geschätzte Fähigkeitsverteilung der Studierenden zusammen mit den Itemschwierigkeiten auf der Logit-Skala dar (Wright Map). Die Itemschwierigkeiten sind zusätzlich nach dem Gerätebereich ihres Items gruppiert. Obwohl sich die Itemschwierigkeiten über das gesamte Fähigkeitspektrum hinweg verteilen, ist ein Mangel von leichten bis mittelschweren Items erkennbar, wodurch die Messpräzision vor allem im unteren Merkmalsbereich eingeschränkt ist.

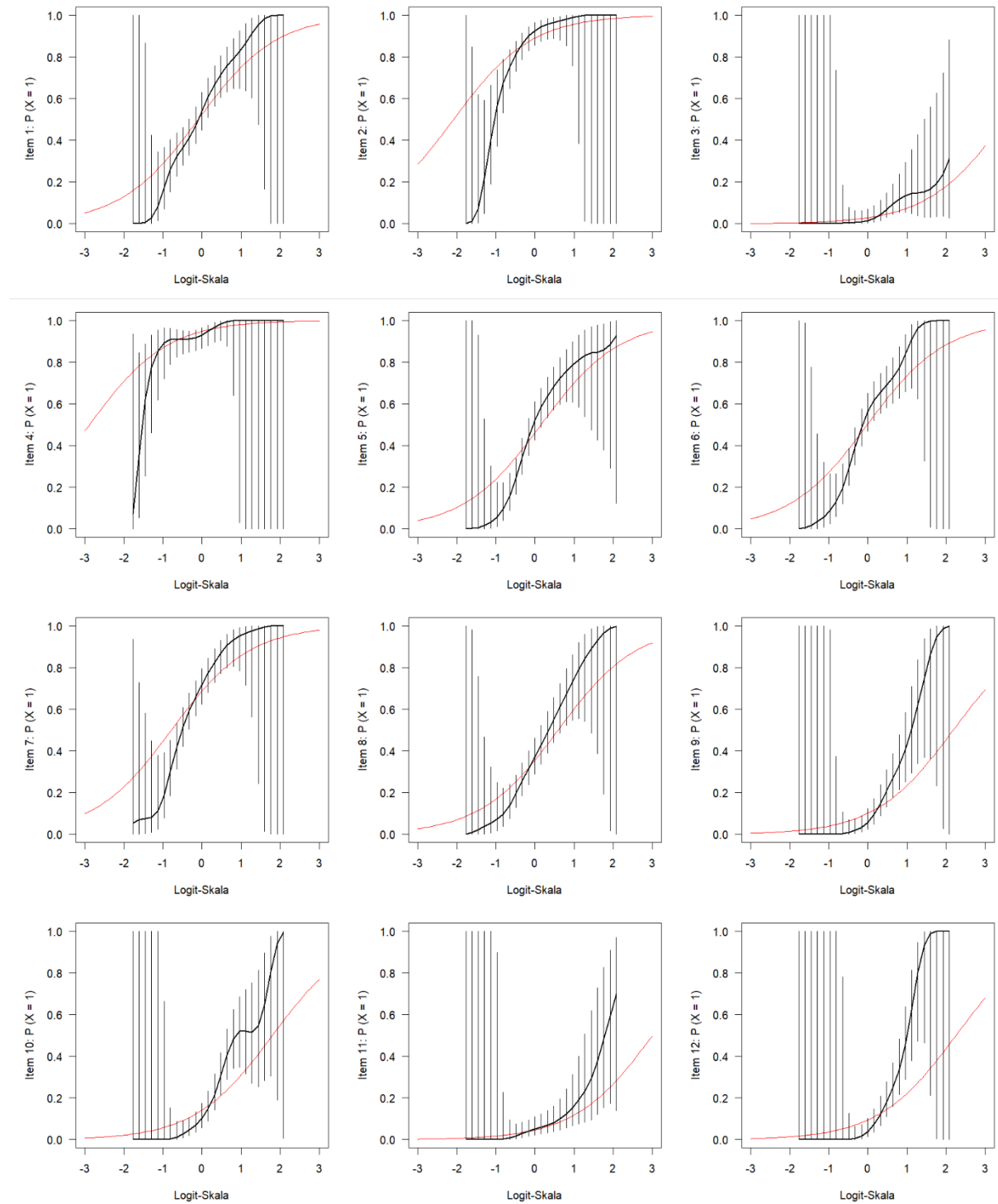
ESM Abbildung 2.2. Verteilung der TPL Steuerungsfähigkeit (links) zusammen mit den Itemschwierigkeitsparametern auf der gemeinsamen Logit-Skala (rechts).



Die Passung des Rasch-Modells auf die Daten lässt sich dennoch insgesamt als gut bewerten. Wie in der ESM Tabelle 2.1 dargestellt, waren die Infit- und Outfitwerte der Items unauffällig und lagen innerhalb der festgelegten Schwellenwerte. Es gab auch nur eine Itempaarung (Item 6 und 7), die eine größere absolute Residualkorrelation als .2 aufwies (Q3-Statistiken: $M = -.00$, $SD = .07$, $Min = -.24$, $Max = .12$), was die Annahme lokaler stochastischer Unabhängigkeit stützt. Die gemeinsame Betrachtung der itemcharakteristischen Kurven und der nicht-parametrischen Antwortfunktionen (ESM Abbildung 2.3) deutete auf keine schwerwiegenden Modellverletzungen hin. Die meisten Items wiesen tendenziell allerdings einen Overfit auf (d. h. eine Tendenz zur Unterschätzung der Erfolgswahrscheinlichkeit für fähige Personen und zur Überschätzung für niedrigfähige Personen).

Elektronisches Supplement 2. TPL-Test und Skalierung

ESM Abbildung 2.3. Modellimplizierte itemcharakteristische Kurven (rot) und nicht-parametrischen Antwortfunktionen (schwarz) auf der Logit-Skala pro Item. Die vertikalen Konfidenzbänder spiegeln die Unsicherheit der Messung wider.



Elektronisches Supplement 2. TPL-Test und Skalierung

Referenzen

- Chen, W.-H. & Thissen, D. (1997). Local Dependence Indexes for Item Pairs Using Item Response Theory. *Journal of Educational and Behavioral Statistics*, 22(3), 265.
<https://doi.org/10.2307/1165285>
- Embretson, S. E. & Reise, S. P. (2013). *Item Response Theory for Psychologists* (Multivariate Applications Series). Mahwah: Erlbaum.
- Partchev, I., Maris, G. & Hattori, T. (2017). *irtoys: A Collection of Functions Related to Item Response Theory (IRT)*. Verfügbar unter: <https://CRAN.R-project.org/package=irtoys>
- R core Team. (2021). *R: A Language and Environment for Statistical Computing*. Vienna Austria: the R Foundation for Statistical Computing. Verfügbar unter: <https://www.r-project.org/>
- Robitzsch, A., Kiefer, T. & Wu, M. (2021). *TAM: Test Analysis Modules*. Verfügbar unter: <https://CRAN.R-project.org/package=TAM>
- Stemmann, J. (2016). *Technische Problemlösekompetenz im Alltag - theoretische Entwicklung und empirische Prüfung des Kompetenzkonstruktes Problemlösen im Umgang mit technischen Geräten*. Dissertation. Universität Duisburg-Essen, Essen. Verfügbar unter: urn:nbn:de:hbz:464-20170118-094621-6
- Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8(3), 370. Verfügbar unter: <https://rasch.org/rmt/rmt83b.htm>