



Linzer Zentrum für
Bildungsforschung
und Evaluation



*Dynamische Messmodelle –
Leistungsvermögen vs. gegenwärtige Leistung
und Leistungsentwicklung*

Christoph Weber

November 2020

Überblick

- Problemaufriss
- Überblick dynamische Messmodelle (DMM)
- Anwendung von DMM auf Lernverlaufsdiagnostik
- Fazit

Problemaufriss

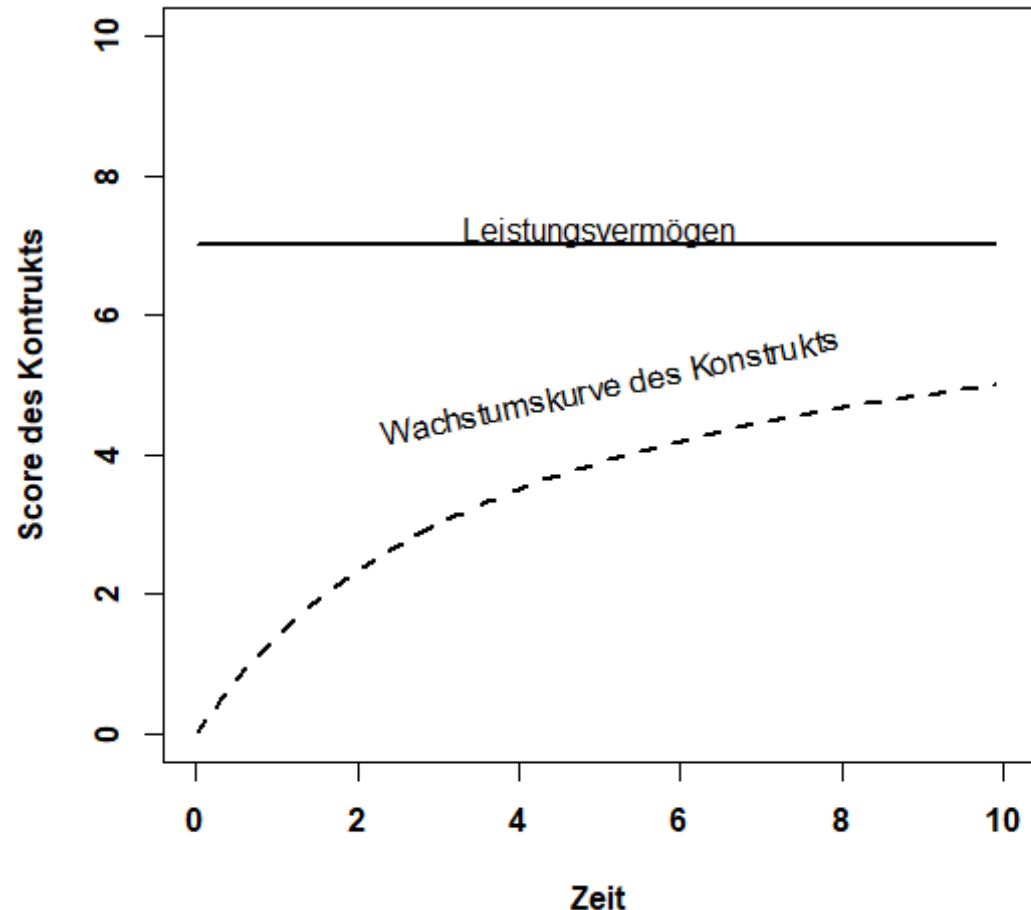
- Probleme querschnittlicher Leistungsfeststellungen
 - Gegenwärtiger Leistungsstand wird mit Leistungsvermögen gleichgesetzt.
 - Grundlage von Schulwahlentscheidungen, ...
 - Einfluss auf Lehrkraft(verhalten) (u.a. Van Houtte et al., 2013; Wang et al., 2018)
- Gegenwärtiger Leistungsstand wird mit Leistungsvermögen gleichgesetzt.
- → Gefahr systematischer Benachteiligung von Gruppen mit (bisher) eingeschränkten Entwicklungschancen (geringerer SES, ...)

Dynamische Messmodelle

(McNeish & Dumas, 2017, 2019; McNeish et al., 2019)

- DMM folgt Grundidee
des *dynamic
assessments*

(u.a. Feuerstein et al., 2002)



Dynamische Messmodelle

(McNeish & Dumas, 2017, 2019; McNeish et al., 2019)

- *Alternative*

*Parametrisierung von
Wachstumsmodellen (Cudeck
& Du Toit, 2002; Preacher & Hancock,
2015)*

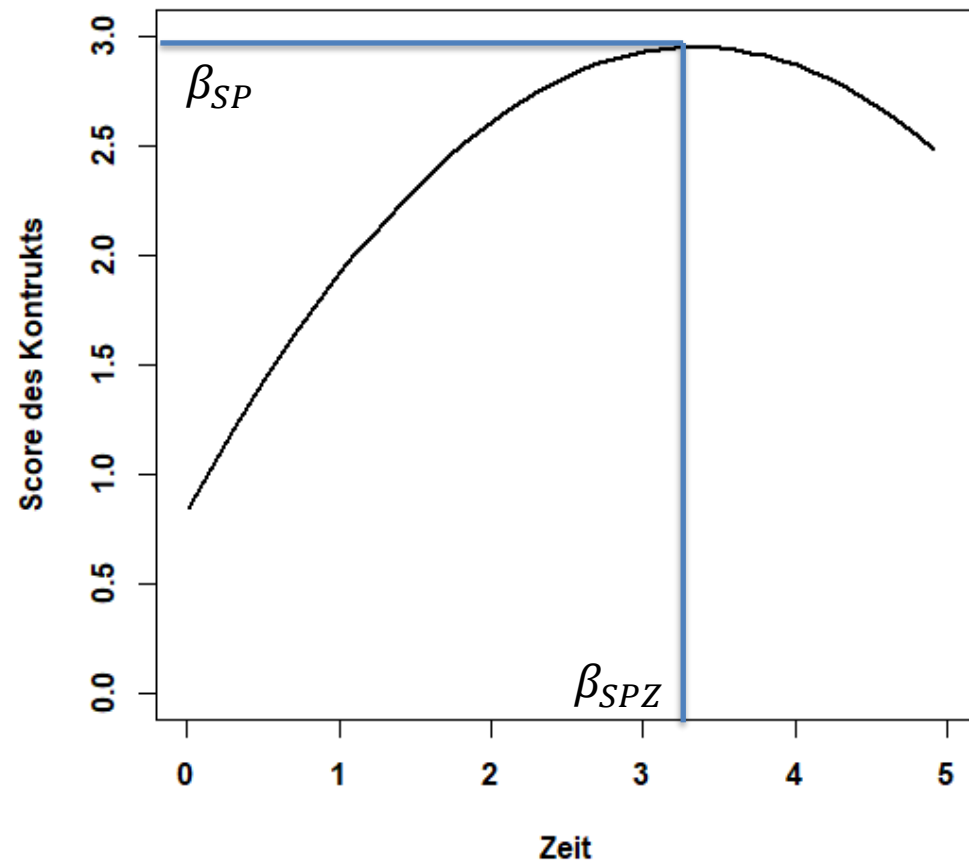
- *Klassische Polynom-
Parametrisierung*

$$\hat{Y} = \beta_0 + \beta_1 \text{Zeit} + \beta_2 \text{Zeit}^2$$

- *Alternative*

Parametrisierung

$$\hat{Y} = \beta_{SP} - (\beta_{SP} + \beta_0) \left(\frac{\text{Zeit}}{\beta_{SPZ}} - 1 \right)^2$$



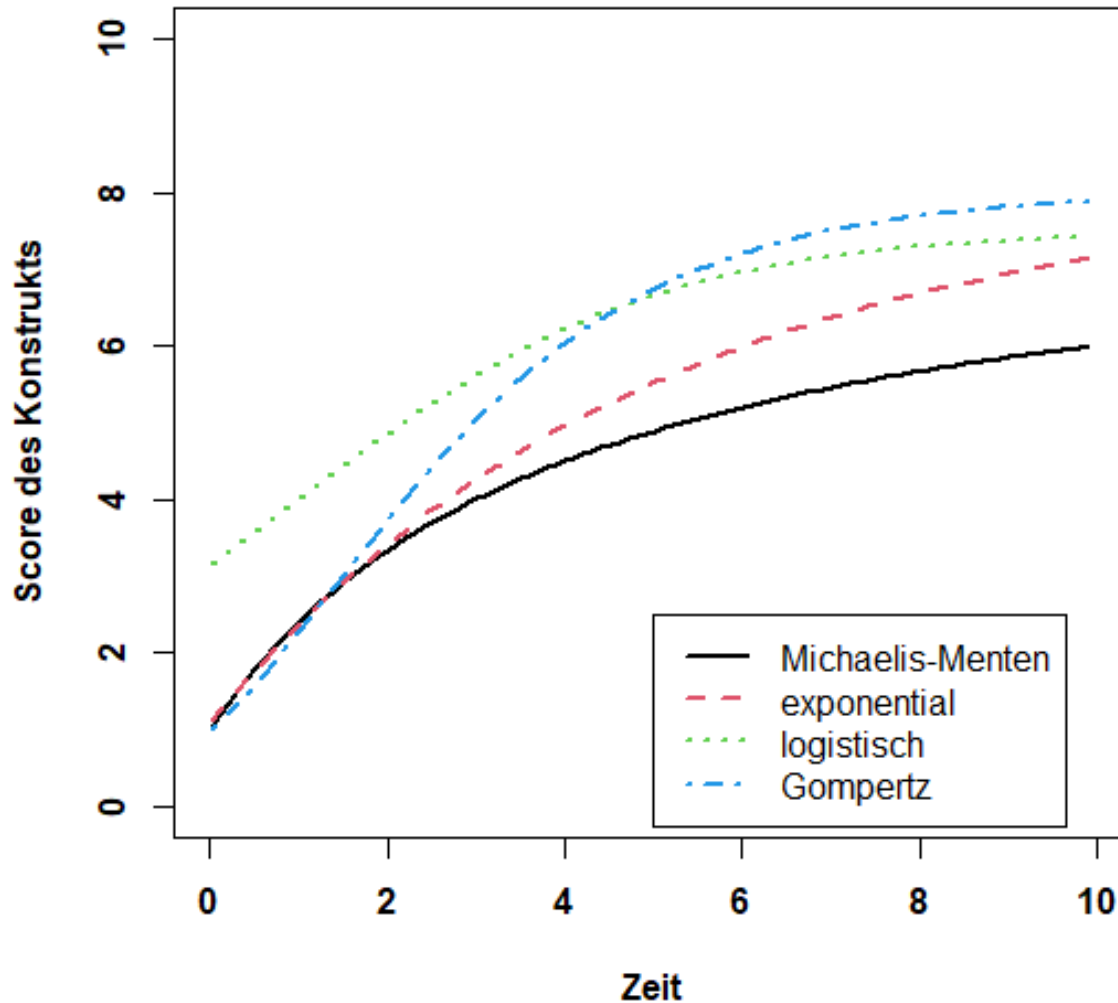
Dynamische Messmodelle

(McNeish & Dumas, 2017, 2019; McNeish et al., 2019)

- DMM-Ansatz greift die Idee des DA und alternativer Parametrisierungen von Wachstumsmodellen auf.
 - Konzeptionalisierung des Leistungsvermögens als obere Asymptote (d.h. Zeit $\rightarrow \infty$)
 - Schätzung der oberen Asymptote als Random Variable in einem nichtlinearen gemischten Modell (SAS proc nlmixed)
- Verschiedene (S- oder J-förmige) Wachstumsfunktionen z.B.
 - Logistische Funktion
$$\beta_{UA} + \frac{(\beta_{OA} - \beta_{UA})}{[1 + e^{-\beta_R(t - \beta_M)}]}$$
 - Exponentielle Funktion
$$\beta_0 + (\beta_{OA} - \beta_0)(1 - e^{t\beta_R})$$
 - Michaelis-Menten
$$\beta_0 + \frac{(\beta_{OA} - \beta_0)t}{[t + \beta_M]}$$
 - Gompertz
$$\beta_{OA} e^{\left[\ln\left(\frac{\beta_0}{\beta_{OA}}\right)e^{(-t\beta_R)}\right]}$$
 - Weitere Funktionen
 - Weibull, Morgan-Mercer-Flodin, ...

Dynamische Messmodelle

(McNeish & Dumas, 2017, 2019; McNeish et al., 2019)



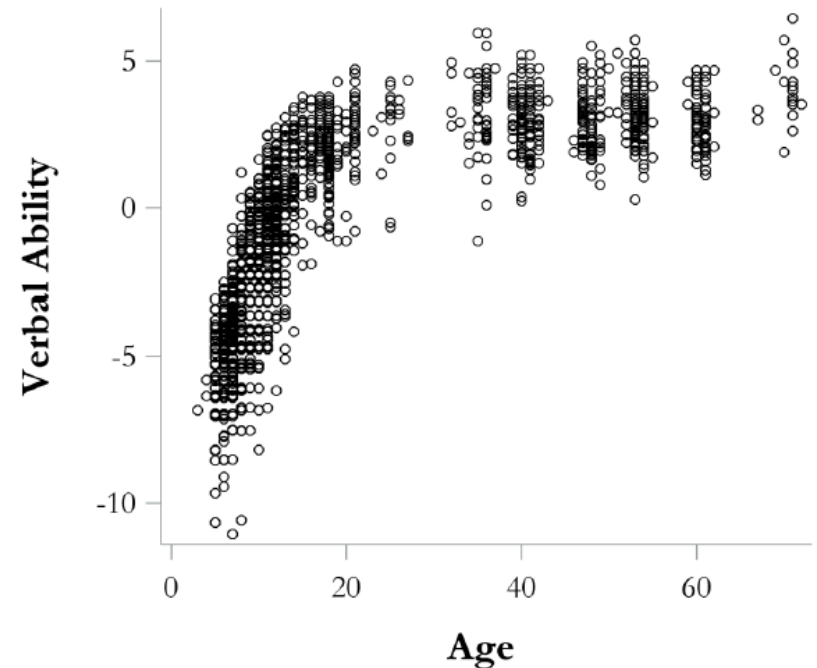
Dynamische Messmodelle

(McNeish & Dumas, 2017, 2019; McNeish et al., 2019)

- Voraussetzungen
 - Werte sind im Längsschnitt auf einer gemeinsamen Skala (z.B. durch vertikale Skalierung, ...; u.a. McArdle et al., 2009; Tong & Kolen, 2007)
- Befunde
 - Höhere prädiktive Validität des Leistungsvermögens im Vergleich zu querschnittlichen Testscores
 - IQ-Leistungsvermögen via DMM (bis 20 Jahre) als Prädiktor späterer IQ-Scores (> 20 Jahre) $R^2 \approx 50\%$, vs. $R^2 \approx 20\%$ bei Querschnittscores als Prädiktor (McNeish et al., 2019).
 - Geringerer Einfluss von SES auf Leistungsvermögen als auf querschnittliche Testscores (Dumas & McNeish, 2017; 2018)
 - SES-Zusammenhänge mit Querschnittscores (Schulleistungen in Mathe und Lesen) Cohen's $f \approx 0.25$, Zusammenhänge mit Leistungsvermögen Cohen's $f = 0.049$.
 - **→ Das Leistungsvermögen von benachteiligten Gruppen (u.a. geringer SES) wird mit Querschnittstests unterschätzt.**

Dynamische Messmodelle (McNeish & Dumas, 2017, 2019; McNeish et al., 2019)

- Bisherige Anwendungsfelder von DMM
- IQ-Längsschnitt (McNeish et al., 2019)
- ECLS-K-Daten (Tourangeau et al., 2009)
 - Lese- und Mathescores für 7 Wellen (Kindergarten bis 8. Schulstufe)
- Administrative Daten (MAP; McNeish & Dumas, 2019)



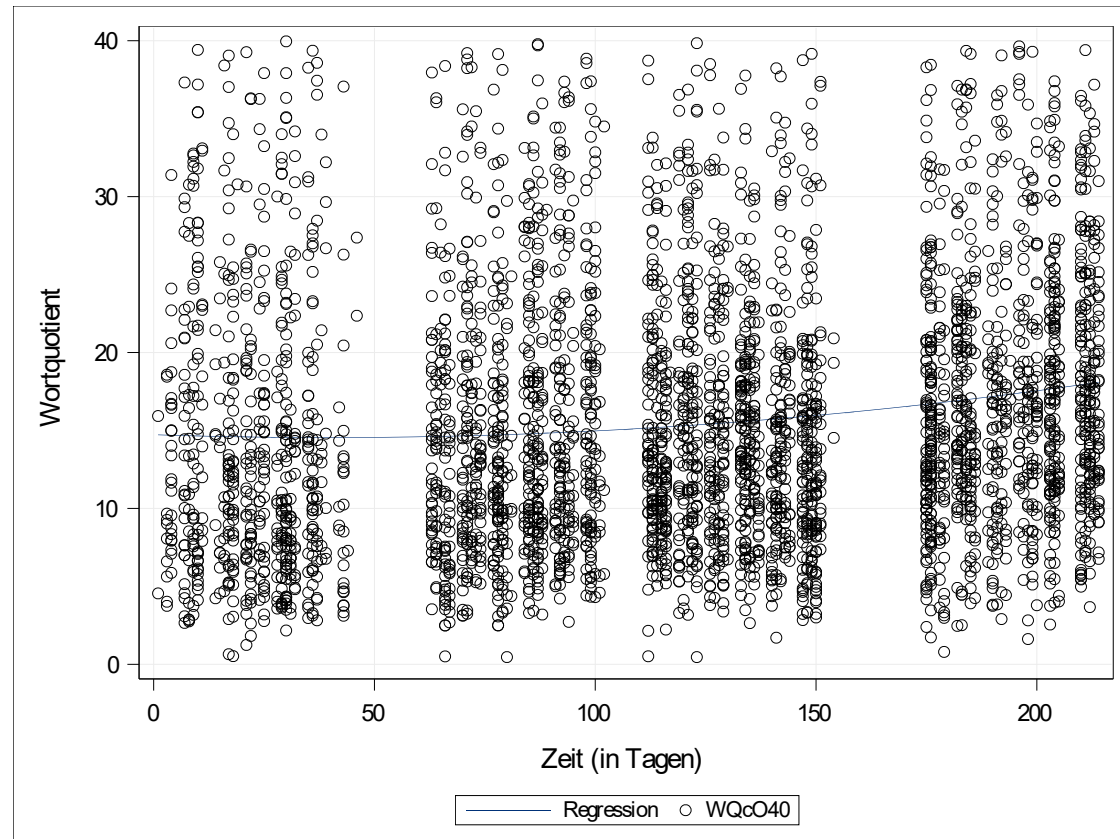
Quelle: McNeish et al. 2019

DMM und Lernverlaufsdiagnostik

- Lernverlaufsdiagnostik (LVD) u.a. Hasselhorn et al., 2014
 - Kurze wiederholte Tests (Paralleltests) in einem Schuljahr zur Sichtbarmachung des Lernverlaufs
- **Datengrundlage**
- Projekt „Wir Wollen`s Wissen!“ des Linzer Zentrums für Bildungsforschung (Weber et al., 2020)
- 28 VS nutzen über 4 Jahre quop (Souvignier et al., 2014)
 - online-basiertes LVD-tool
 - Kurze, selbstadministrierte Einzeltests (ca. 10 -15 Minuten)
 - 8 Messzeitpunkte pro Schuljahre
 - Lesen und Mathe
- **Fragestellung**
- Können DMM für LVD-Daten genutzt werden?
 - Konkret für die Wortleseflüssigkeit im Verlauf der 1. Klasse (n=518)
- Ergibt sich eine bessere Vorhersage des Leseverständnisses (Ende 2. Klasse, ELFE II, Lenhardt et al., 2017) durch das geschätzte Leistungsvermögen als durch Querschnittscores?

Voranalysen

- Vergleich möglicher Wachstumsfunktionen
 - MSE auf Basis SAS proc nlin
 - Gompertz und Exponential weisen geringsten MSE (69.3) auf
 - Andere Modelle häufig Konvergenzprobleme



MSE lineares Modell (68.3)

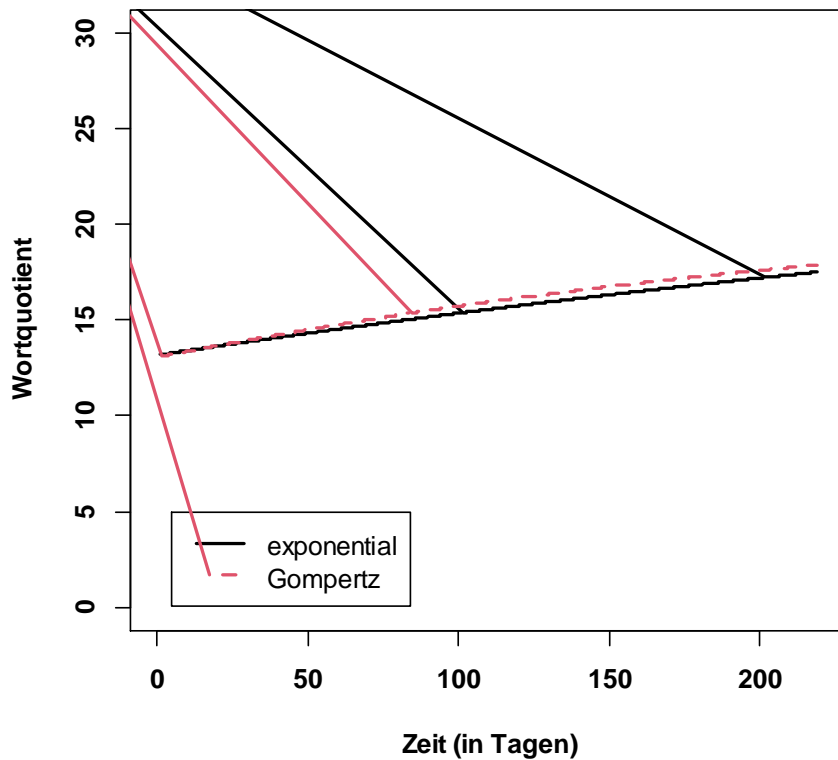
nlmixed Ergebnisse

	Gompertz	Exponential
Fixe Effekte	Schätzer (SE)	Schätzer (SE)
<i>Intercept</i>	13.072*** (0.438)	13.173*** (0.410)
<i>Obere Asymptote (Leistungsvermögen)</i>	20.943*** (1.854)	26.526*** (6.341)
<i>Wachstumsrate</i>	0.005** (0.001)	0.002 (0.001)
Random Effekt Varianzen		
<i>Intercept</i>	63.767*** (0.201)	64.928*** (0.190)
<i>Obere Asymptote (Leistungsvermögen)</i>	257.404** (31.01)	760.640 (228.648)
Random Effekt Korrelation		
<i>Intercept, obere Asymptote</i>	-0.566*** (0.087)	-0.659*** (0.070)
Model fit		
<i>BIC</i>	26200	26200

Anmerkungen: *** $p < .001$, ** $p < .01$, * $p < .05$

nlmixed Ergebnisse

- Prädiktive Validität



Pearsonsche Korrelationskoeffizienten
Anzahl der Beobachtungen

	ELFE Wort_T	ELFE Satz_T	ELFE Text_T	ELFE Gesamt_T
Wort- quotient t6	0.29519 290	0.31549 290	0.20016 290	0.29782 290
Wort- quotient t7	0.48632 291	0.47247 291	0.35711 291	0.48116 291
Wort- quotient t8	0.47306 259	0.49442 259	0.40702 259	0.50817 259
Predicted Leistungs- vermögen	0.43999 329	0.47656 329	0.40502 329	0.48411 329

Zusammenfassung und Fazit

- DMM stellen interessante Alternative zur längsschnittlichen Modellierung von Leistungsdaten dar
- LVD-Daten über ein Jahr scheinen weniger geeignet zu sein
 - Zu kurzer Zeitraum (kein monoton-abschwächendes Wachstum) → Vertikale Skalierung über 4. Jahre VS?
 - LVD geringe Reliabilität/Validität (Kurztests, hohe Ratewahrscheinlichkeit, z.T. sehr hohe Streuung innerhalb von Schüler/innen im Zeitverlauf, viele Extremwerte, Rolle des Lehrkraftverhalten, selbstadministrierte Tests bei 1. Klässler/innen, ...)
 - Skaleneffekte? (negative Intercept-Asymptotenkorrelation, negative Korrelation von t1-Werten mit t8-Werten)
 - Annahme gleicher Skala (KTT) durch Paralleltests? → alternative Skalierungsverfahren (IRT)



Linzer Zentrum für
Bildungsforschung
und Evaluation



Danke für die
Aufmerksamkeit!

Literatur

- Feuerstein, R., Feuerstein, R. S., Falik, L. H., & Rand, Y. A. (2002). The dynamic assessment of cognitive modifiability: The Learning Propensity Assessment Device: Theory, instruments and techniques, Rev. and exp. ed. Of The dynamic assessment of retarded performers. ICELP Publications.
- Cudeck, R., & Du Toit, S. H. (2002). A version of quadratic regression with interpretable parameters. *Multivariate Behavioral Research*, 37(4), 501-519.
- Dumas, D. G., & McNeish, D. M. (2017). Dynamic measurement modeling: Using nonlinear growth models to estimate student learning capacity. *Educational Researcher*, 46(6), 284-292.
- Hasselhorn, M., Schneider, W. & Trautwein, U. (Hrsg.). (2014). *Lernverlaufsdagnostik (Tests und Trends, Bd. 12, 1. Aufl.)*. Göttingen, Niedersachs: Hogrefe Verlag.
- McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological methods*, 14(2), 126.
- McNeish, D., & Dumas, D. (2017). Nonlinear growth models as measurement models: A second-order growth curve model for measuring potential. *Multivariate behavioral research*, 52(1), 61-85.
- McNeish, D., & Dumas, D. G. (2019). Scoring repeated standardized tests to estimate capacity, not just current ability. *Policy Insights from the Behavioral and Brain Sciences*, 6(2), 218-224.
- McNeish, D., Dumas, D. G., & Grimm, K. J. (2019). Estimating new quantities from longitudinal test scores to improve forecasts of future performance. *Multivariate Behavioral Research*, 1-16.
- Preacher, K. J., & Hancock, G. R. (2015). Meaningful aspects of change as novel random coefficients: A general method for reparameterizing longitudinal models. *Psychological methods*, 20(1), 84.
- Souvignier, E., Förster, N. & Salaschek, M. (2014). quop: Ein Ansatz internetbasierter Lernverlaufsdagnostik mit Testkonzepten für Lesen und Mathematik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdagnostik (Tests und Trends, Bd. 12, 1. Aufl., S. 239–256)*. Göttingen, Niedersachs: Hogrefe Verlag.
- Tong, Y., & Kolen, M. J. (2007). Comparisons of methodologies and results in vertical scaling for educational achievement tests. *Applied Measurement in Education*, 20(2), 227-253.
- Tourangeau, K., Nord, C., Lê, T., Sorongon, A. G., & Najarian, M. (2009). *Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K): Combined User's Manual for the ECLS-K Eighth-Grade and K-8 Full Sample Data Files and Electronic Codebooks*. NCES 2009-004. National Center for Education Statistics.
- Van Houtte, M., Demanet, J., & Stevens, P. A. (2013). Curriculum tracking and teacher evaluations of individual students: selection, adjustment or labeling?. *Social Psychology of Education*, 16(3), 329-352.
- Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, 24(3-5), 124-179.
- Weber, C., Kemethofer, D. & Helm, C. (2020). Wir Wollen's Wissen! - Wie es um die Kompetenzen unserer Schüler/-innen steht - Diagnostik und Fortschrittsmonitoring als Ausgangslage für Schul- und Unterrichtsentwicklung an Schulen in herausfordernder Lage. *SchulVerwaltung Spezial*, 22(4), 75-76.