

Zeitvorgabeproblematik bei speeded Powertests am Beispiel LPS-2

Kreuzpointner, L. & Hristova, S.

Institut für Psychologie, Universität Regensburg

Einführung

Das Leistungsprüfsystem 2 (LPS-2, Kreuzpointner, Lukesch & Horn, im Druck) misst die kognitive Leistungsfähigkeit basierend auf der 3-Ebenen-Theorie von John Carroll (1993; 2005). Die elf Subtests (1: Allgemeinwissen, 2: Anagramme, 3: Figurenfolgen, 4: Zahlenfolgen, 5: Buchstabenfolgen, 6: Mentale Rotation, 7: Flächenzahl, 8: Linienmuster, 9: 8. Zeichen, 10: Zeilenvergleich, 11: Addieren) operationalisieren dabei Faktoren erster Ordnung, die zu vier Faktoren zweiter Ordnung (Gc: Kristalline Intelligenz, Gf: Fluide Intelligenz, Gv: Visuelle Wahrnehmung, Gs: Kognitive Schnelligkeit) zusammengefasst werden können. Alle Testleistungen zusammen dienen der Schätzung der allgemeinen kognitiven Leistungsfähigkeit als Faktor dritter Ordnung. Während die Relevanz der Bearbeitungsgeschwindigkeit in den Subtests 6 bis 11 recht eindeutig ist, stellen wir im Folgenden eine Untersuchung zum Einfluss der Geschwindigkeit auf die Leistungseinordnung der Subtests 1 bis 5 vor.

Design

Um den Effekt, den eine echte Powertesting auf die Testleistung hat, zu operationalisieren, wurde den Probanden nach der regulären Testung des gesamten LPS-2 zusätzlich Zeit für die Bearbeitung und Korrektur der ersten fünf Subtests eingeräumt, die diese mit einem andersfarbigen Stift durchführten. Zur Effektbestimmung wurde neben dem Vergleich der zeitbegrenzten mit der zeitlich unreglementierten Leistung mit dem Mehrfachwahl-Wortschatz-Intelligenztest (MWT-A; Lehrl, 1977) und dem Bochumer Matrizenest (BOMAT; Hossiep, Truck & Hasella, 1999) zwei inhaltsähnliche externe Kriterien erhoben.

Hypothesen & Methoden

Es wurde erwartet, dass die Nachbearbeitung des LPS-2 zu höheren Rohwerten führt, die Reihenfolge der Probanden hinsichtlich der Leistung aber relativ konstant bleibt, die Leistungen also hoch korrelieren. Zudem sollte sich auch der Zusammenhang der Tests untereinander und mit den anderen beiden Tests kaum unterscheiden. Die Mittelwertdifferenzen werden mittels Cohens d' standardisiert. Bivariate Zusammenhänge werden mit der Produkt-Moment-Korrelation berechnet. Zur Überprüfung der Äquivalenz der Zusammenhangsstrukturen dienen Strukturgleichungsmodelle.

Ergebnisse

Im Mittel werden sieben bis neun Aufgaben in der nach der regulären Testung zur Verfügung gestellten Zeit richtig gelöst, wobei die Streuung in den beiden sprachlichen Tests deutlich größer und der Effekt der zweiten Bearbeitung damit deutlich geringer ist als in den Tests zum logischen Denken (Tabelle 1). Die Testleistungen korrelieren zwischen .61 und .77 und somit etwas unter der Marke, um von statistisch äquivalenten Tests sprechen zu können. Die erhöhten Korrelation der Differenzen mit den Leistungen mit Zeitbegrenzung zeigen, dass bei den Subtests 1 und 2 (-.51 und -.65) für Probanden mit niedrigeren regulären Leistungen größere Verbesserungen möglich gewesen waren. In den generell als am schwierigsten geltenden Subtests 4 und 5 zeigt sich dieser Effekt hingegen nicht (-.08 bzw. .04, vgl. Abbildung 1). Die Korrelationen der LPS-2 Leistungen mit MWT und BOMAT unterscheiden sich in den beiden Bedingungen kaum (Tabelle 2). Die stärksten Abweichungen in kongruenten Beziehungen zeigen sich für Subtest 2 und MWT (.30 zu .43) sowie Subtest 3 und BOMAT (.30 zu .05). In den durch die Zusammenfassung reliableren Ergebnissen der Dimensionen nivellieren sich diese Unterschiede jedoch zunehmend. Hinsichtlich der Zusammenhangsstruktur der LPS-2 Subtests zeigt sich durch die zeitlich unbegrenzte Bearbeitungsmöglichkeit eine Verschiebung der Wichtigkeit der fluiden Intelligenz für die Gewichtung von G von .80 zu .91 (Abbildung 2). Aufgrund der geringen Stichprobengröße ist der Modellfit der beiden SEM - anders als in mehreren anderen Untersuchungen - schlecht (mit Zeitbegrenzung: $\chi^2=82.1$, $df=47$; RMSEA=.11; ohne Zeitbegrenzung: $\chi^2=98.1$, $df=47$; RMSEA=.13). Die beta-Gewichte der Subtests ändern sich hingegen wenig, lediglich für die Subtests 4 und 5 sind Verringerungen zu beobachten.

Diskussion

In den sprachlichen Subtests 1 und 2 sind Deckeneffekte beobachtbar. Probanden, die innerhalb der Zeit schon sehr viele Aufgaben richtig beantwortet hatten, konnte sich kaum mehr verbessern, während langsameren Probanden diese Möglichkeit noch offenstand. In den formallogischen Subtests 3, 4 und 5 zeigen sich Verbesserungen über das gesamte Leistungsspektrum und sehr wenige Probanden mit nur minimaler Verbesserung. Es wird deutlich, dass die Bearbeitungsgeschwindigkeit nur einen kleinen Teil der Leistungsvarianz ausmacht und Probanden, die in der regulären Bedingung gut sind, sich bei Zeitzugabe ebenso verbessern wie Probanden, die generell langsamer arbeiten. Es zeigt sich aber auch, dass es einzelne Probanden gibt, die nur ohne zeitliche Begrenzung ihr volles Leistungsrepertoire abrufen können. Wird fluide Intelligenz ohne zeitliche Begrenzung erhoben, zeigen die resultierenden Werte einen größeren Einfluss auf die Schätzung des G-Faktors, als wenn die Leistungen unter Zeitdruck erbracht werden müssen. Hierin könnte eine Erklärung für die immer wieder unterschiedlichen Ergebnisse hinsichtlich des Einflusses der fluiden Intelligenz auf die Schätzung der allgemeinen kognitiven Leistungsfähigkeit liegen (mal moderat, mal hoch), die zu den unterschiedlichen Ausdifferenzierungen der hierarchischen Intelligenzmodelle geführt hat, insbesondere die Diskussion zwischen Verfechtern des Cattell/Horn-Modells (Horn & Blankson, 2005) und des Carroll-Modells (2005; vgl. aber auch Gustafsson, 1984).

Literatur

- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment* (pp. 69-76). New York: Guilford.
- Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179-203.
- Horn, J. L. & Blankson, N. (2005). Foundations for better understanding cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment* (pp. 41-68). New York: The Guilford Press.
- Hossiep, R., Truck, D. & Hasella, M. (1999). *Bochumer Matrizenest (BOMAT)*. Göttingen: Hogrefe.
- Kreuzpointner, L., Lukesch, H. & Horn, W. (im Druck). *Leistungsprüfsystem 2 (LPS-2)*. Göttingen: Hogrefe.
- Lehrl, S. (1977). *Mehrfachwahl-Wortschatz-Intelligenztest (MWT-A)*. Göttingen: Hogrefe.

Tabelle 1: Mittlere Differenzen, Standardabweichungen und Effektstärken des Unterschiede der Leistungen in den Subtests 1-5 und den Dimensionen G_c und G_f des LPS-2 in den Bedingungen mit Zeitbegrenzung und ohne Zeitbegrenzung sowie die Korrelationen der Ergebnisse in den beiden Bedingungen (r_{m_o}) und die Korrelation der Differenz mit der Leistung in der zeitbegrenzten Version ($r_{m_{diff}}$).

	M	SD	d'	r_{m_o}	$r_{m_{diff}}$
1	8.32	6.09	1.37	.77	-.51
2	7.66	5.36	1.43	.74	-.65
3	8.35	2.75	3.04	.69	-.38
4	9.42	4.09	2.30	.61	-.08
5	8.92	3.83	2.33	.66	.04
G _c	15.98	9.71	1.65	.79	-.56
G _f	26.69	7.59	3.52	.71	-.09

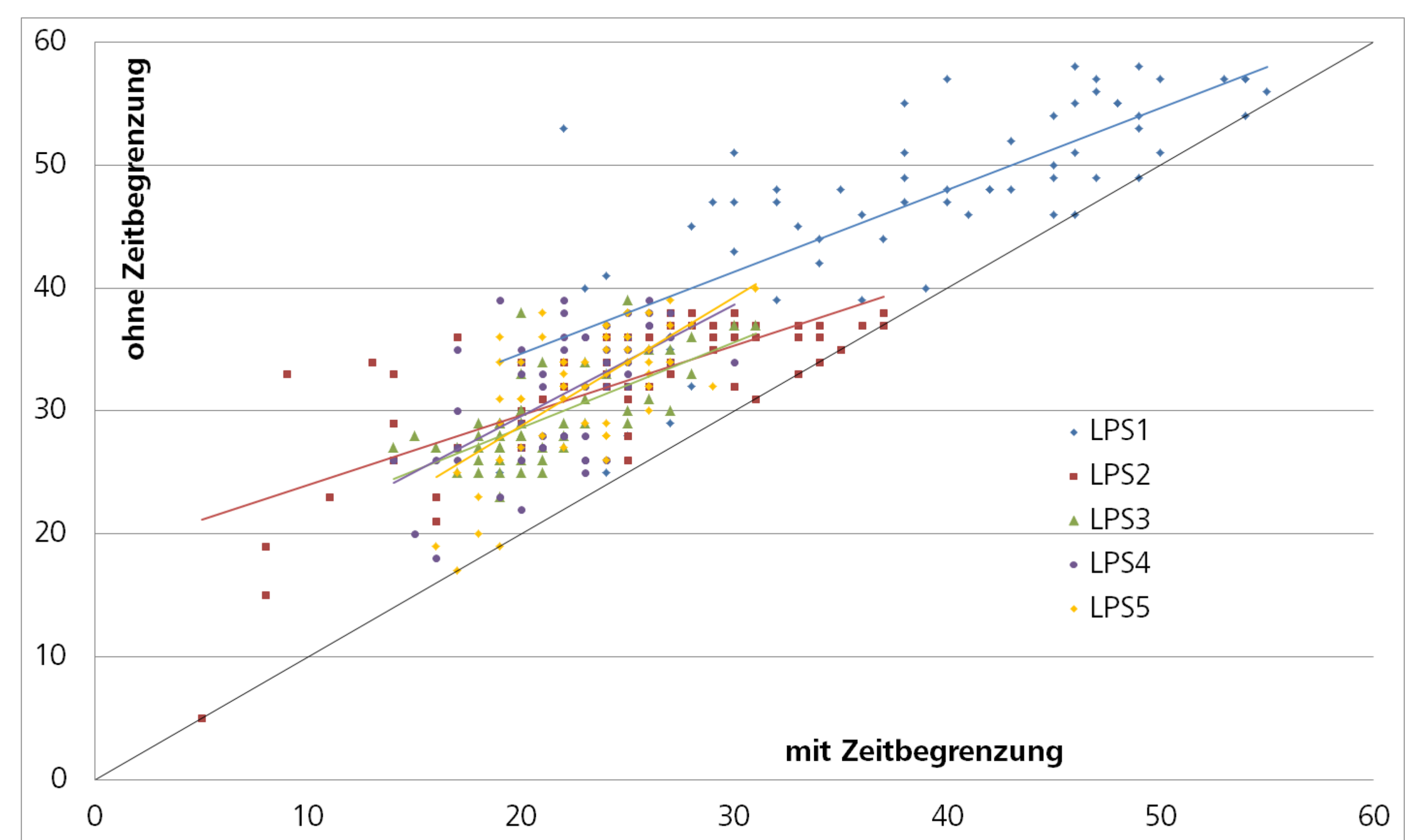


Abbildung 1: Streudiagramme und Trendgeraden der LPS-2 Subtests 1-5 der Leistungen mit Zeitbegrenzung gegen die Leistungen ohne Zeitbegrenzung

Tabelle 2: Produkt-Moment-Korrelationen der Subtests 1-5, der Dimensionen G_c und G_f und des Gesamtwertes des LPS-2 mit dem Mehrfachwahl-Wortschatztest und dem Bochumer Matrizenest sowie entsprechende Mittelwerte und Standardabweichungen der 62 Probanden

	Mit Zeitbegrenzung				Ohne Zeitbegrenzung			
	MWT	BOMAT	M	SD	MWT	BOMAT	M	SD
1	.64	.20	39.08	9.41	.65	.20	47.40	8.16
2	.30	.22	24.65	7.99	.43	.23	32.31	6.10
3	.11	.30	20.90	3.46	.10	.05	29.26	3.50
4	-.14	.32	22.02	3.45	.00	.37	31.44	5.14
5	.14	.38	23.10	3.25	.14	.41	32.02	5.12
G _c	.53	.23	63.73	15.69	.61	.23	79.71	13.00
G _f	.04	.40	66.02	8.36	.10	.39	92.71	10.77
G	.25	.45	286.73	38.01	.27	.46	329.40	37.09
M	30.32	15.32						
SD	3.69	3.54						

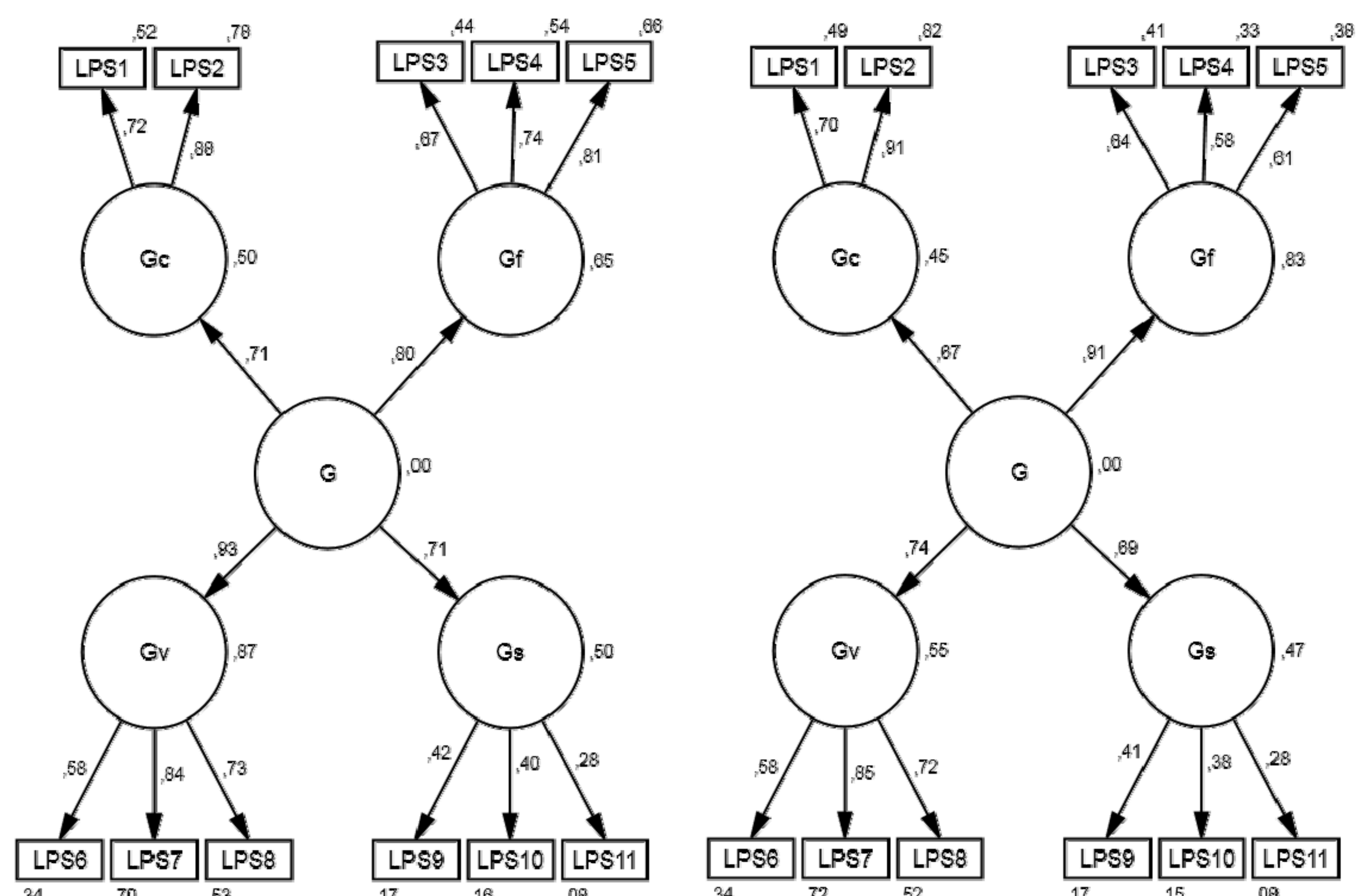


Abbildung 2: SEM des LPS-2 links für die Standardversion mit Zeitbegrenzung, rechts für die Ergebnisse ohne Zeitbegrenzung. Aufgrund der geringen Stichprobengröße (n=62) ist der globale Modellfit für beide Modelle schlecht.