

Item Parceling: Bildung von Testteilen oder Item-Päckchen

1 Ausgangssituation

Bei der Analyse von Strukturgleichungsmodellen ist es oftmals nötig, aus verschiedenen Gründen die Anzahl der vorhandenen Variablen zu reduzieren. Sofern Variablen nicht einfach weggelassen werden können, geschieht dies in der Regel, indem mehrere Items aufaddiert (oder gemittelt) werden und diese Summen- (bzw. Mittelwerts-)variablen dann statt der einzelnen Items als Indikatoren im Strukturgleichungsmodell verwendet werden.

Dieses sog. Item Parceling, d.h. das Bilden von Item-Päckchen, wird seit einigen Jahren zunehmend häufiger eingesetzt (vgl. Bandalos, 2002; Bandalos & Finney, 2001; Little, Cunningham, Shahar & Widaman, 2002; Marsh, Hau, Balla & Grayson, 1998). Die Popularität dieses Verfahrens liegt vor allem darin, dass verschiedene Probleme durch das Item Parceling behoben oder zumindest abgemildert werden können (Bandalos & Finney, 2001). Häufig ist die Stichprobe zu klein, um sehr große Modelle mit vielen Parametern zu analysieren. Als Daumenregel werden oftmals 5 bis 10 Personen pro zu schätzendem Parameter veranschlagt (vgl. Bentler & Chou, 1987); bei 30 Parametern müsste man folglich 150 bis 300 Personen untersucht haben. Sind die manifesten Variablen nicht normalverteilt, können verzerrte Schätzungen und nicht korrekte Gütemaße für den Modellfit resultieren. Desweiteren können die manifesten Variablen mehrdimensional sein, indem z. B. Messfehlerkorrelationen auftreten können.

- Wenn die Stichprobe relativ klein ist, so ist es oftmals nötig, die Anzahl der zu schätzenden Parameter zu reduzieren, damit stabile Parameterschätzungen ermöglicht werden und die Modellgüte korrekt geschätzt werden kann (Bagozzi & Heatherton, 1994; Bagozzi & Edwards, 1998; Bandalos & Finney, 2001; Little et al., 2002; MacCallum, Widaman, Zhang & Hong, 1999). Wird z. B. ein Modell mit zwei latenten Variablen, die jeweils durch 9 Items gemessen werden, analysiert, so müssten 37 Parameter geschätzt werden (18 Fehlervarianzen, 16 Faktorladungen, 2 Varianzen und 1 Kovarianz der latenten Variablen). Wird die Anzahl der Indikatoren durch Item Parceling auf drei Indikatoren pro latentem Konstrukt reduziert, so würden dann nur noch 13 Parameter zu schätzen sein (6 Fehlervarianzen, 4 Faktorladungen, 2 Varianzen und 1 Kovarianz der latenten Variablen). Für ein solches Modell würden nunmehr wesentlich weniger Personen benötigt.

- Sind die einzelnen Items grobstufig kategorial, so fallen daraus gebildete Itempäckchen als Indikatoren feiner abgestuft aus, mit einer gleichmäßigeren, einer Normalverteilung ähnlicheren Verteilung. Damit kann die Modellgüte zuverlässiger bestimmt werden als im Fall der Verwendung der einzelnen Items (Bandalos, 2002). Der gleiche Vorteil ergibt sich, wenn Items durch Boden- oder Deckeneffekte der Skalen (extrem leichte bzw. extrem schwierige Items) schief verteilt sind und Items mit jeweils entgegengerichteter Schiefe zu Päckchen zusammengefaßt werden (so daß die Verteilung dann eher symmetrisch ausfällt).
- Die meisten Studien zum Item Parceling befassen sich mit dem Problem der Dimensionalität von Item Parcels. Die bisherige Forschung hat gezeigt, dass die Bildung von Item Parcels vor allem dann Vorteile hat, wenn die gemessenen Variablen eindimensional sind. Liegt dagegen keine Eindimensionalität vor – z. B. durch Methodeneffekte –, so gibt es keine einheitliche Empfehlung, wie man konkret vorgehen sollte. Häufig werden solche Items, die mehr als eine Dimension messen, über die Päckchen verteilt, so dass alle Dimensionen in jedem Päckchen vorhanden sind. Dadurch soll sich ein verbesserter Modellfit ergeben (vgl. Kishton & Widaman, 1994; Little et al., 2002). Es besteht jedoch auch das Problem, dass durch diese Strategie wesentliche Informationen unterdrückt werden können (Bandalos, 2002; Hall, Snell & Singer Foust, 1999).

2 Durchführung

Es gibt verschiedene Arten der Bildung von Item-Päckchen, die je nach Bedarf und Fragestellung gewählt werden können:

1. Bildung von Item-Päckchen nach Faktorladungen, Mittelwert und Varianz

Um möglichst ähnliche Item-Päckchen zu konstruieren, kann man die Faktorladungen, Mittelwerte und Varianzen der Items heranziehen. In einer CFA auf der Basis von Items werden zunächst die Faktorladungen der Items bestimmt, zusätzlich werden deskriptive Statistiken berechnet. Es werden dann solche Items auf die verschiedenen Päckchen (meist drei) verteilt, die sich möglichst ähnlich sind. Somit sollten hinterher alle Päckchen bezüglich der statistischen Kriterien vergleichbar sein. Es könnte u. U. aber auch sinnvoll sein, alle Items mit einer hohen Faktorladung dem ersten Päckchen zuzuordnen, welches dann das latente Konstrukt am besten misst, und die Items mit niedrigeren Faktorladungen auf die anderen Päckchen zu verteilen. Durch die so ermöglichte Gewichtung unterschiedlich gut messender Items kann die Teststärke erhöht werden.

2. Bildung von Item-Päckchen nach Inhalten

Messen verschiedene Items jeweils einen unterschiedlichen Aspekt eines gemeinsamen Merkmals, z. B. Verhalten oder Kognitionen, objektive oder subjektive Maße, etc., so können die Päckchen nach inhaltlichen Kriterien zusammengestellt werden. Somit würden z. B. ein Päckchen mit Verhaltensitems und ein Päckchen mit Kognitionsitems gebildet werden. Bei sehr homogenen Konstrukten bietet sich diese Strategie jedoch nicht an.

3. Bildung von Item-Päckchen nach der Verteilung der Variablen

Sind die Items nicht normal-, sondern schiefverteilt, so können Items mit jeweils betragsmäßig vergleichbarer, aber entgegengesetzter Richtung der Schiefe zu Päckchen zusammengefaßt werden. Durch Items mit entgegengerichteter (also jeweils positiver und negativer) Schiefe würde für die Päckchen dann eine eher normalverteilte Summenvariable resultieren. Dies ist jedoch nur sinnvoll, wenn die Nichtnormalität der Items durch Meßfehler zustande gekommen ist (wie im Falle von Boden- und Deckeneffekten bei einzelnen sehr leichten bzw. sehr schwierigen Items). Sind die Items dagegen aus inhaltlichen Gründen alle gleichermaßen schief verteilt (weil das zugrundeliegende Konstrukt, das gemessen werden soll, selbst nicht normalverteilt ist), so kann Päckchenbildung dies nicht ändern (und sollte es auch nicht, da andernfalls wesentliche Information über die Verteilungsform des Konstrukts verlorengehen würde).

3 Voraussetzungen

Vor der Bildung von Item-Päckchen sollte folgendes geprüft werden:

1. Überprüfung der Dimensionalität der Konstrukte

Bei Mehrdimensionalität könnten separate Itempäckchen für jede Dimension gebildet werden, bei Eindimensionalität würden sich dagegen statistische Kriterien für die Itempäckchenbildung anbieten.

2. Überprüfung der Verteilung der Items

Wenn viele Items nicht normalverteilt sind, kann durch eine geeignete Strategie die Verteilung der Item-Päckchen normaler werden. Mehr Items per Parcel sind besser als wenige Items, falls dadurch die Eindimensionalität der Parcels erhalten bleiben kann.

Literatur

- Bagozzi, R. P. & Edwards, J. R. (1998). A general approach for representing constructs in organizational research. *Organizational Research Methods, 1*, 45–87.
- Bagozzi, R. P. & Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: Application to state self-esteem. *Structural Equation Modeling, 1*, 35–67.
- Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling, 9*, 78–102.
- Bandalos, D. L. & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides & R. E. Schumacker (Hg.), *Advanced structural equation modeling: New developments and techniques* (S. 269–296). Mahwah, NJ: Lawrence Erlbaum.
- Bentler, P. M. & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods and Research, 16*, 78–117.
- Hall, R. J., Snell, A. F. & Singer Foust, M. (1999). Item parceling strategies in SEM: Investigating the subtle effects of unmodeled secondary constructs. *Organizational Research Methods, 2*, 233–256.
- Kishton, J. M. & Widaman, K. F. (1994). Unidimensional versus domain representative parceling of questionnaire items: An empirical example. *Educational and Psychological Measurement, 54*, 757–765.
- Little, T. D., Cunningham, W. A., Shahar, G. & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9*, 151–173.
- MacCallum, R. C., Widaman, K. F., Zhang, S. & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*, 84–99.
- Marsh, H. W., Hau, K.-T., Balla, J. R. & Grayson, D. (1998). Is more ever too much? the number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*, 181–220.