

Clemens Hetschko\*, Louisa von Reumont\*\* und Ronnie Schöb\*\*\*

## Die Schwierigkeit, soziale Wohlfahrt zu messen: Einbettungseffekte im OECD Better Life Index

Um die soziale Wohlfahrt anhand verschiedener Indikatoren, die das Wohlbefinden der Menschen beeinflussen, zu messen, muss man klären, wie wichtig uns die einzelnen Indikatoren sind, d.h. mit welchem Gewicht sie in die Wohlfahrtsfunktion eingehen sollten. Die OECD lädt mit dem Better Life Index alle Menschen ein, selber zu entscheiden, wie wichtig ihnen die einzelnen Indikatoren sind. Unser Experiment zeigt, dass die daraus resultierenden Gewichtungen kein zuverlässiges Maß für die wahren Vorlieben der Menschen für einzelne Indikatoren sind. Ursächlich hierfür sind Einbettungseffekte, die die Validität des Better Life Index als Instrument für die Ermittlung einer Wohlfahrtsfunktion grundsätzlich infrage stellen.

### WOHLFAHRTSMESSUNG UND DER OECD BETTER LIFE INDEX

Um die Wohlfahrt über das Bruttoinlandsprodukt (BIP) hinaus zu messen, schlug die vom ehemaligen französischen Präsidenten Nicolas Sarkozy eingesetzte Stiglitz-Sen-Fitoussi-Kommission ein breit angelegtes, multidimensionales statistisches System zur Messung des Wohlbefindens der Menschen vor (vgl. Stiglitz et al. 2009, S. 12). Nur eine kohärente und multidimensionale Datenbasis, die sowohl objektive als auch subjektive Wohlbefindensmaße umfasst, ermögliche es Politikern, die Lebensqualität der Bürger richtig zu messen und zu verbessern. Um eine solche Datenbasis zu generieren, muss sich die Gesellschaft auf eine Auswahl von Indikatoren verständigen. Zudem muss jeder Indikator gewichtet werden, um die relative Bedeutung der einzelnen Indikatoren für die Gesamtwohlfahrt zu bestimmen. Stiglitz et al. (2009, S. 65) konstatieren für die bekannten Wohlfahrtsmaße, wie z.B. den *Human Development Index* (HDI), dass die Gewichtung der einzelnen Indikatoren zwar transparent sei, jedoch zugleich willkürliche und nicht normativ fundierte Gewichtungen erzeuge. Dieser Mangel ist äußerst problematisch, zum einen angesichts der Wirkung, den Wohlfahrtsmaße in der öffentlichen Diskussion entfalten und zum anderen, weil aus ihnen Politikprioritäten abgeleitet werden.

\* Dr. Clemens Hetschko ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Finanzwissenschaft mit dem Schwerpunkt Internationale Finanzpolitik an der Freien Universität Berlin.

\*\* Louisa von Reumont ist Senior Strategy Consultant bei Jung von Matt/SPORTS GmbH.

\*\*\*Prof. Dr. Ronnie Schöb ist Inhaber des Lehrstuhls für Finanzwissenschaft mit dem Schwerpunkt Internationale Finanzpolitik an der Freien Universität Berlin.

Die OECD hat diese Kritik aufgegriffen und ein viel beachtetes Tool entwickelt, um das Gewichtungsproblem zu lösen, den *OECD Better Life Index* (im Folgenden: BLI). Mit ihm können alle Bürger weltweit die relative Bedeutung verschiedener objektiver und subjektiver Wohlbefindensindikatoren selber bestimmen. In einer internetbasierten Umfrage können Menschen selbst festlegen, wie wichtig ihnen elf vorgegebene Wohlfahrtsdimensionen (Wohnverhältnisse, Einkommen, Beschäftigung, Gemeinsinn, Bildung, Umwelt, Zivilengagement, Gesundheit, Lebenszufriedenheit, Sicherheit und Work-Life-Balance) sind, die wiederum insgesamt untergeordnete 24 Einzelindikatoren umfassen. Daraus errechnen sich die Gewichte, mit denen diese Dimensionen in die individuelle Wohlfahrtsfunktion eingehen (<http://www.oecdbetterlifeindex.org>). Der BLI hat einige mediale Aufmerksamkeit erhalten und spielt eine wichtige Rolle in der aktuellen wissenschaftlichen Debatte zur Wohlfahrtsmessung (vgl. z.B. Durand 2015; Decancq 2017; Marković et al. 2016; Balestra et al. 2018). Auch in die Politik fand er bereits Eingang. So hat die österreichische Regierung eine modifizierte Version des BLI verwendet, um die Lebensqualität ihrer Bürger zu messen und mit der anderer Länder zu vergleichen (vgl. WIFO 2012).

Der BLI ist einfach, transparent und demokratisch; damit erscheint er auch normativ gut begründet. Dennoch zeigt unser Experiment, dass der BLI nicht in der Lage ist, zu einer konsistenten Gewichtung unterschiedlicher Wohlfahrtsindikatoren zu kommen. Der Index unterliegt Einbettungseffekten (*embedding effects*), die dazu führen, dass je nachdem wie einzelne wohlfahrtsrelevante Indikatoren in übergeordneten

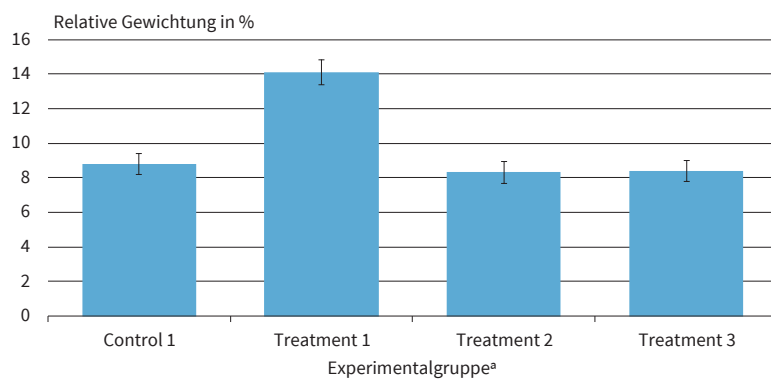
Dimensionen für die Befragung zusammengefasst werden, völlig unterschiedliche Gewichtungen der Indikatoren resultieren. Der BLI ist demzufolge kein geeignetes Maß für die Messung von Wohlfahrt. Ferner stellt sich die Frage, inwieweit es grundsätzlich möglich ist, durch Befragungen zu konsistenten Gewichtungen wohlfahrtsrelevanter Faktoren, die unser Leben beeinflussen, zu gelangen (vgl. Hetschko, von Reumont und Schöb 2019).

### EINBETTUNGSEFFEKTE IM BETTER LIFE INDEX: EXPERIMENTELLE BEFUNDE

Einbettungseffekte sind von der Bewertung öffentlicher Güter durch Befragungen bekannt. Der Wert, der einem bestimmten Projekt beigemessen wird, variiert stark je nachdem, ob das Projekt für sich allein genommen oder als Teil eines übergeordneten Programms mit mehreren Projekten bewertet wird. Bei der Einzelbewertung bemessen die Befragten dem konkreten Projekt systematisch einen höheren Wert zu (vgl. Kahneman und Knetsch 1992). Ähnliches vermuteten wir im Hinblick auf den BLI, da die Teilnehmer 24 Wohlfahrtsindikatoren nicht einzeln gewichten, sondern als Teil von elf übergreifenden Dimensionen (vgl. z.B. OECD 2015). So sind beispielsweise die Indikatoren Arbeitslosigkeit (gemessen durch die Beschäftigungsquote und die Langzeitarbeitslosenquote), durchschnittliches Arbeitsentgelt und Arbeitsplatzsicherheit in der Dimension Beschäftigung zusammengefasst. Dies macht die Onlinebefragung auch transparent: Fährt man mit der Maus über den Dimensionstitel, erscheinen die eingebetteten Indikatoren in einem Pop-Up-Fenster. Klickt man den Titel an, erhält man detaillierte Informationen zu den eingebetteten Indikatoren. Im Idealfall sollte es keinen Unterschied machen, ob die Befragten die Indikatoren einzeln oder als Teil einer Dimension gewichten bzw. welche anderen Indikatoren noch Teil der Dimension sind.

Um das zu überprüfen und den BLI auf potenzielle Einbettungseffekte zu testen, haben wir Studierende neun deutscher Universitäten gebeten, Bewertungen für die einzelnen Dimensionen des BLI vorzunehmen (für eine detaillierte Beschreibung vgl. Hetschko, von Reumont und Schöb 2019). Dabei haben wir in einer ansonsten genauen Replikation des Original-BLI (im Folgenden »RBLI«) den in die Dimension Beschäftigung eingebetteten Indikatorensetz verändert. Bei einer ersten Treatmentgruppe (»Treatment 1«) haben wir die Dimension Beschäftigung in zwei Dimensio-

Abb. 1  
Experimentelle Befunde



<sup>a</sup> Für Control 1 wie auch für Treatment 2 und 3 zeigen die Balken die durchschnittlichen relativen Gewichtungen der Dimension Beschäftigung. Der Balken Treatment 1 steht für die Summe der durchschnittlichen relativen Gewichtungen der Dimensionen Jobqualität und Arbeitsmarkt. 95%-Konfidenzintervalle werden durch Whiskers dargestellt.

Quelle: Hetschko, von Reumont und Schöb (2019).

© ifo Institut

nen, Arbeitsmarkt und Jobqualität, zerlegt, wobei die neue Dimension Arbeitsmarkt die Indikatoren Beschäftigungsquote und Langzeitarbeitslosenquote beinhaltet und sich die neue Dimension Jobqualität aus den Indikatoren Arbeitsentgelt und Arbeitsplatzsicherheit zusammensetzt. Wenn es keine Einbettungseffekte gibt, müssten diese beiden neuen Dimensionen in der Summe dieselbe relative Gewichtung erhalten wie die Originaldimension Beschäftigung, da die für die Wohlfahrtsmessung relevanten Indikatoren in beiden Fällen dieselben sind. Die Bewertung der originalen Dimension Beschäftigung überließen wir der Kontrollgruppe »Control 1«. Wie Abbildung 1 jedoch zeigt, hat die Treatmentgruppe 1, die die beiden Dimensionen Jobqualität und Arbeitsmarkt separat gewichtet hat, den vier Indikatoren ein weitaus höheres Gewicht für die Wohlfahrt beigemessen als die Kontrollgruppe 1, die nur eine Dimension mit allen vier Indikatoren gewichtet hat.

### SYMBOLHAFTIGKEIT VON DIMENSIONEN UND INDIVIDUELLE AUFFASSUNGEN

Eine wesentliche Ursache für den gefundenen Einbettungseffekt mag darin liegen, dass die Befragten nicht die spezifischen Indikatoren einer Dimension gewichten, sondern eher ihre vom Titel der Dimension ausgelöste persönliche Auffassung vom Inhalt der Dimension. Sobald andere Dimensionstitel verwendet werden, ändern sich deren Bewertungen, ungeachtet dessen, dass die zugrunde liegenden Indikatoren die gleichen sind. Um die Stichhaltigkeit dieser Erklärung zu überprüfen, haben wir ebenfalls untersucht, ob eine Dimension dasselbe relative Gewicht erhält, wenn sie nur bestimmte Indikatoren-Untergruppen enthält. Bei der »Treatment 2«-Version des RBLI beinhaltet die Beschäftigungsdimension ausschließlich jobspezifische Aspekte, d. h. Arbeitsentgelt und Arbeitsplatzsicherheit. Befragte in »Treatment 3« gewichteten mit der Dimension Beschäftigung nur die

Arbeitsmarktindikatoren Langzeitarbeitslosenquote und Beschäftigungsquote. Mindestens eine der beiden Gruppen hätte die Dimension Beschäftigung niedriger gewichten müssen als die Kontrollgruppe, wenn Einbettungseffekte keine Rolle gespielt hätten. Wie wir in Abbildung 1 sehen, erhielt die Dimension mit der Bezeichnung »Beschäftigung« jedoch stets die gleiche relative Gewichtung. Die Bewertung wurde komplett von der mit dem Dimensionstitel verbundenen vorgefertigten Auffassung dominiert; die spezifischen zugrunde gelegten Indikatoren blieben hingegen völlig unbeachtet. Das liegt jedoch nicht daran, dass sie nicht ausreichend sichtbar gewesen wären: Wir finden die gleichen Ergebnisse in Untergruppen von Befragten, die sich der Indikatoren eindeutig bewusst gewesen sein dürften, da sie sich die entsprechenden Zusatzinformationen aufgerufen oder sich für den Gewichtungsvorgang relativ lange Zeit gelassen haben (vgl. Hetschko, von Reumont und Schöb 2019).

### IMPLIKATIONEN FÜR DIE VALIDITÄT DES BETTER LIFE INDEX UND ANDERER UMFRAGEBASIERTER WOHLFAHRTSMASSE

Unsere Ergebnisse stellen die Eignung des BLI als Maß für Präferenzen von Menschen für Indikatoren der Lebensqualität grundlegend infrage. Der Einbettungseffekt bringt es mit sich, dass wir das wahre Gewicht eines Indikators in den Augen eines Befragten nicht kennen. Der Wert, den er dem Indikator zuordnet, reflektiert Aspekte, die der Indikator nicht misst, und lässt gegebenenfalls Aspekte außen vor, die der Indikator widerspiegelt. Ein Vergleich von Indikatoren, wie z.B. die Aussage »Arbeitsentgelt ist den Menschen wichtiger als Umweltqualität«, kann aus den Ergebnissen des *Better Life Index* ebenso wenig abgeleitet werden wie Vergleiche von Gewichtungen zwischen Ländern. Da vorgefasste Ansichten über Dimensionen dem kulturellen Umfeld und regionalen Erfahrungen entspringen können, kann es sein, dass Angehörige zweier Gesellschaften, die die vier originären Indikatoren der Dimension Beschäftigung gleichermaßen einschätzen, am Ende aber trotzdem zu unterschiedlichen relativen Gewichtungen kommen. Es nützt auch nichts, die Veränderung der relativen Gewichtungen im Zeitverlauf zu vergleichen. Wenn für Menschen »Beschäftigung« zunehmend wichtiger wird, könnte das an einem Aspekt der Beschäftigung liegen, der nicht unter den eingebetteten Indikatoren aufgeführt ist, wie etwa weniger Zeitdruck. Der BLI würde fälschlicherweise implizieren, dass Arbeitslosigkeit, Jobsicherheit und Arbeitsentgelt wichtiger geworden seien.

Es gibt ein weiteres Problem, das wie die zuvor genannten Probleme nicht nur den BLI betrifft, sondern potenziell jede umfragebasierte Methode der Wohlfahrtsmessung, die eine Einbettungsstruktur vorgibt, um die Komplexität des Gewichtungsverfahrens zu reduzieren. Sie könnten Politikentschei-

den dazu verhelfen, das Wohlfahrtsmaß im Sinne einer von ihnen verfolgten Agenda zu manipulieren. So würde ein arbeitnehmerfreundlicher Politiker Arbeitsplatzsicherheit und Arbeitsentgelt jeweils einzeln als eigenständige Dimension abfragen wollen, wohingegen ein arbeitgeberfreundlicher Politiker eher dafür sorgen würde, dass beide Aspekte ein und derselben Dimension angehören.

### KRITISIEREN IST EINFACH, DOCH WIE KANN MAN ES BESSER MACHEN?

Unsere Ergebnisse identifizieren ein Problem, präsentieren aber keine Lösung. Die ist auch nicht leicht zu finden. Es gibt einige mögliche Lösungsansätze, wie z.B. den von Benjamin et al. (2014). Sie präsentieren Befragten hypothetische diskrete Auswahlsszenarien, bei denen kleine Veränderungen zweier unterschiedlicher Lebensqualitätsaspekte verglichen werden müssen. Solche Umfragen bedienen sich einer weit aus vorsichtigeren Herangehensweise als der BLI. Sie können zwar keine globale Wohlfahrtsfunktion ableiten, aber immerhin lokale Eigenschaften der Wohlfahrtsfunktion im Status quo. Entsprechend nützlich können sie für Evaluationen von konkreten, klar abgegrenzten Politikreformen sein. Für die Bewertung umfassenderer, fundamentaler Reformen oder den Vergleich von Ländern sind sie hingegen nicht geeignet. Bescheidenheit ist daher angesagt. Die Fokussierung auf kleinere Verbesserungen der sozialen Wohlfahrt erscheint uns ein vielversprechenderer Forschungsansatz zu sein als die Suche nach einem globalen Wohlfahrtsindex (vgl. Hetschko, von Reumont und Schöb 2019).

### LITERATUR

- Balestra, C., R. Boarini und E. Tosetto (2018), »What Matters Most to People? Evidence from the OECD Better Life Index Users' Responses«, *Social Indicators Research* 136(3), 907–930.
- Benjamin, D. J., O. Heffetz, M. Kimball und N. Szembrot (2014), »Beyond Happiness and Satisfaction: Toward Well-Being Indices Based on Stated Preference«, *American Economic Review* 104(9), 2698–2735.
- Decancq, K. (2017), »Measuring Multidimensional Inequality in the OECD Member Countries with a Distribution-Sensitive Better«, *Social Indicators Research* 131(3), 1057–1086.
- Durand, M. (2015), »The OECD Better Life Initiative: How's life? And the Measurement of Well-being«, *Review of Income and Wealth* 61(1), 4–17.
- Hetschko, C., L. von Reumont und R. Schöb (2019), »Embedding as a Pitfall for Survey-based Welfare Indicators: Evidence from an Experiment«, *Journal of the Royal Statistical Society Series A* 182(2), 517–539.
- Kahneman, D. und J. L. Knetsch (1992), »Valuing Public-Goods – the Purchase of Moral Satisfaction«, *Journal of Environmental Economics and Management* 22(1), 57–70.
- Marković, M., S. Zdravković, M. Mitrović und A. Radojičić (2016), »An Iterative Multivariate Post Hoc I-Distance Approach in Evaluating OECD Better Life Index«, *Social Indicators Research* 126(1), 1–19.
- OECD (2015), *How's Life? 2015: Measuring Well-Being*, OECD Publishing, Paris.
- Stiglitz, J. E., A. K. Sen und J.-P. Fitoussi (2009), *Report by the Commission on the Measurement of Economic Performance and Social Progress*, verfügbar unter: <https://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report>.
- WIFO (2012), *Mehr als Wachstum Messung von Wohlstand und Lebensqualität in ausgewählten OECD-Ländern Ausgangslage und Aufgabenstellung*, Wien, mimeo.