

GUTACHTEN ZUR METHODIK DES SHOPPINGCENTER PERFORMANCE REPORT DEUTSCHLAND 2023 (ECOSTRA GMBH)

Prof. Dr. Björn Christensen und Prof. Dr. Manuel Stegemann
(Fachhochschule Kiel / Hochschule Bielefeld) im Auftrag des
German Council of Shopping Places e. V.



WARUM GIBT ES DIESES NEUE GUTACHTEN?

Sehr geehrte Damen und Herren,

fünf Jahre nach dem ersten Gutachten zur Methodik des Shoppingcenter Performance Report galt es noch einmal genau hinzuschauen, um zu überprüfen, ob die Autoren, Herausgeber und Partner aus den Hinweisen aus dem Gutachten zum SCPR 2018 etwas gelernt hatten. Warum? Weil es immer noch Marktteilnehmer und Medien gibt, die auf der Basis des Rankings einen Erfolg oder Misserfolg des Shoppingcenter-Standortes bestimmen können wollen. Wir verstehen unsere Verantwortung als der Verband, der alle Handelsimmobilien-Klassen und die dazugehörigen Akteure aus Handel, Gastronomie und vielschichtigen Dienstleistern unter seinem Dach vereint, auch darin, dass mit qualifizierten Argumenten und wissenschaftlichen Untersuchungen solche Mieterbefragungen einsortiert werden als das, was sie sind und leisten können.

Wir schließen uns daher den Aussagen der beiden Gutachter an und zitieren aus dem Fazit gerne vorab diese drei Absätze:

Die Validität und Reliabilität des SCRP 2023 ist aus wissenschaftlich-methodischer Sicht als relativ gering einzustufen. Dies ist in erster Linie auf die Performance-Messung sowie die Stichprobengröße zurückzuführen. Des Weiteren kann nicht ausgeschlossen werden, dass es aufgrund der Stichprobenauswahl zu Verzerrungen hinsichtlich der Repräsentativität kommt.

Der Single-Item-Ansatz zur Messung der Performance ist nicht empfehlenswert. Die Uneindeutigkeit der Frage sowie die hohe Komplexität erschweren die intersubjektive Vergleichbarkeit der Bewertungen.

Die insgesamt leicht verbesserte Gesamtqualität des SCPR 2023 sollte jedoch nicht darüber hinwegtäuschen, dass das Ranking und die Mittelwerte der Center wissenschaftlich nicht als valide zu bewerten sind und aus Sicht der Gutachter nicht als Grundlage für weitreichende (De-)Investitionsentscheidungen Anwendung finden sollten.

Der Erfolg eines gut integrierten und aus Sicht der Mieter und Eigentümer sowie der finanzierenden Bank bzw. Partner auch wirtschaftlich funktionierenden Shopping Centers hat viele große und kleine Faktoren im Blick. Die Nähe des Centers am Kunden als Dienstleister, der hohe Grad an Kundenorientierung der Einzelhandels- und Gastronomie Mietern und deren Waren und Dienstleistungen sowie das vielseitige Angebot an zum Beispiel Arztpraxen, Kindergärten bzw. öffentlichen Nutzungen, gepaart mit einer für alle funktionierenden Infrastruktur zur bequemen Erreichbarkeit, sind in unterschiedlicher Ausprägung Grundlagen jedes erfolgreichen Standortes.

Eines gilt aber für alle: Nur gemeinsam können alle Akteure des Handels, der Handelsimmobilienbranche und der Stadt für lebendige Innenstädte sorgen. Mehr Verständnis für einander hilft, und am Ende gilt das alte Sprichwort: „Man schreibt sich auseinander, man spricht sich zusammen.“

Nutzen Sie unser Gutachten, um mit Ihren Partnern ins Gespräch zu kommen.



Ingmar Behrens



Ingmar Behrens,
Generalsekretär
des German Council of Shopping Places e. V.

Gutachten zur Methodik des Shoppingcenter Performance Report Deutschland 2023

INHALTSVERZEICHNIS

Zielsetzung dieses Gutachtens	4
Kurzbeschreibung des Shoppingcenter Performance Reports Deutschland 2023	4
Methodik zur Erhebung der Performance von Shoppingcentern	5
Die Auswahl der Stichprobe	6
Die Stichprobengröße pro Shoppingcenter	6
Darstellung und Interpretation der Ergebnisse	8
Fazit zur Methodik des Shoppingcenter Performance Reports Deutschland 2023	8
Quellen	8
Rückblick: Gutachten zum SCPR 2018	11-19

ZIELSETZUNG DIESES GUTACHTENS

Die Methodik des Shoppingcenter Performance Reports (SCPR) 2023 (im Folgenden zitiert als ecostra 2023) soll im Rahmen dieses Gutachtens bezüglich der zugrundeliegenden Methodik wissenschaftlich bewertet werden. Dies erfolgt insbesondere hinsichtlich des Anspruchs seitens der Herausgeber, dass der SCPR ein wichtiges Instrument zur Darstellung der wirtschaftlichen Performance sowie der Leistungsfähigkeit von 400 Shoppingcentern sein soll und wertvolle Dienste für Standortexpansion und Filialnetzoptimierung leisten soll (vgl. S. 1 ff. der Vorbemerkungen in ecostra 2023).

Ergänzend stellt dieses Gutachten die methodischen Anpassungen des SCPR im Vergleich zum Jahr 2018 heraus und bewertet diese im Hinblick auf die Erhebungsqualität und Aussagekraft des Reports.

Hinweis: Da wesentliche methodische Vorgehensweisen seit der letzten Stellungnahme der Autoren zum SCPR 2018 unverändert geblieben sind, werden viele Aussagen und Textbausteine auch in der aktualisierten Stellungnahme beibehalten. Sie behalten ihre Gültigkeit.

KURZBESCHREIBUNG DES SHOPPINGCENTER PERFORMANCE REPORTS DEUTSCHLAND 2023

Der Report, dessen Hauptbestandteil ein Ranking von 400 Shoppingcentern in Deutschland bildet, wird seit dem Jahr 2011 jährlich von der ecostra GmbH, einer Wirtschafts-, Standort- und Strategieberatung, erstellt und kostenpflichtig vertrieben. Weiterhin begleiten die Fachzeitschriften Immobilien Zeitung und TextilWirtschaft als Medienpartner den jährlichen Report.

Als Ziel des Reports gibt die ecostra GmbH an, den Marktteilnehmern systematisch und übersichtlich Informationen zur wirtschaftlichen Performance von Shoppingcentern aus Mietersicht bereitzustellen. Diese Informationen könnten nach Eigenangabe eine wichtige Grundlage für Managemententscheidungen wie beispielsweise Expansion oder Filialnetzoptimierung darstellen (vgl. S. 1 ff. der Vorbemerkungen in ecostra 2023).

»Da wesentliche methodische Vorgehensweisen seit der letzten Stellungnahme der Autoren zum SCPR 2018 unverändert geblieben sind, werden viele Aussagen und Textbausteine auch in der aktualisierten Stellungnahme beibehalten.«



Die Datengrundlage für den Report bildet eine Online-Umfrage unter Ansprechpartnern von Unternehmen, die als Mieter Stores in deutschen Shoppingcentern betreiben. Die Umfrage wurde von ecostra im Zeitraum vom 01.06.2023 bis zum 20.08.2023 durchgeführt. Lediglich 94 vollständig ausgefüllte Fragebögen dienen als Datengrundlage des Reports sowie des Rankings der 400 Shoppingcenter (vgl. S. 1 ecostra 2023).

METHODIK ZUR ERHEBUNG DER PERFORMANCE VON SHOPPINGCENTERN

Die verwendeten Items in einem Fragebogen sind entscheidend für die Qualität der Daten und somit auch für die Aussagekraft. Im SCPR 2023 wurde, wie in den Vorjahren auch, die Performance eines Stores in einem Shoppingcenter anhand einer einzigen Frage gemessen (vgl. S. 12 ecostra 2023). In der wissenschaftlichen Literatur wird ein sog. Single-Item-Measurement nur in Ausnahmefällen befürwortet, wenn das zu messende Merkmal sehr konkret und homogen, d. h. wenig facettenreich, ist (vgl. Kreis et al. 2021, S. 102 ff.; Diamantopoulos et al. 2012). Die Performance eines Shoppingcenters ist jedoch zweifellos ein sehr heterogenes und komplexes Merkmal mit verschiedenen Facetten. Es ist daher auf Basis wissenschaftlicher Qualitätskriterien nicht empfehlenswert, dieses Merkmal mit nur einer einzigen Frage zu messen. Zahlreiche Studien zeigen, dass in einem solchen Fall die Messung sowohl in Bezug auf die Reliabilität (Zuverlässigkeit) als auch auf die Validität (Gültigkeit) eingeschränkt ist. Daher sind die Ergebnisse nur mit großer Vorsicht zu interpretieren (vgl. z. B. Diamantopoulos et al. 2012; Sarstedt, M. und Wilczynski, P. 2009).

Aus der Messung der Performance als Single-Item ergibt sich ein weiteres Problem: Die verwendete Frage zur Messung der Performance ist kompliziert formuliert und erfordert eine abwägende Bewertung verschiedener Faktoren durch die Teilnehmenden. Es soll die Umsatzleistung eines einzelnen Stores in Relation zu den Standortkosten und gleichzeitig relativ zu anderen Stores des Unternehmens in weiteren Shoppingcentern bewertet werden. Eine derart komplexe Frage, die verschiedene Faktoren verknüpft, wird in der Regel von Teilnehmenden unterschiedlich verstanden. Dies wiederum führt zu intersubjektiv verschiedenen Bewertungsdimensionen der Teilnehmenden. Die Eindeutigkeit einer Fragestellung ist jedoch erforderlich, um eine gemeinsame Verständnisbasis zur Performance zu schaffen und um die Ergebnisse vergleichbar zu machen (vgl. z. B. Moosbrugger, H. und Kelava, A. 2012, S. 65 f.).

Darüber hinaus beinhaltet die Frage einen relativen Vergleich zu den anderen Stores des jeweiligen Unternehmens. Dies ist methodisch als problematisch anzusehen, da jedes teilnehmende Unternehmen eine andere Vergleichsbasis insbesondere hinsichtlich der individuellen Standortkosten hat, was die Vergleichbarkeit der Bewertungen erheblich einschränkt. Für den Report ist eine hohe Vergleichbarkeit jedoch von besonderer Bedeutung, da die Shoppingcenter auf Basis der Durchschnittsbewertung mit zwei Nachkommastellen gerankt werden.

»Lediglich 94 vollständig ausgefüllte Fragebögen dienen als Datengrundlage des Reports sowie des Rankings der 400 Shoppingcenter.«

»Zahlreiche Studien zeigen, dass in einem solchen Fall die Messung sowohl in Bezug auf die Reliabilität (Zuverlässigkeit) als auch auf die Validität (Gültigkeit) eingeschränkt ist. Daher sind die Ergebnisse nur mit großer Vorsicht zu interpretieren.«

Die verwendete Bewertungsskala ist angelehnt an ein Schulnotensystem (1 = „sehr gut“ bis 5 = „mangelhaft“), der Inhalt der Frage bezieht sich jedoch auf die monetäre Situation der einzelnen Stores im Vergleich zu unternehmensinternen vergleichbaren Stores in Shoppingcentern. Eine Übertragung in ein Schulnotensystem ist bei Befragungen zwar nicht unüblich, allerdings unterliegt diese Art der Bewertung dem Einfluss der „individuellen Strenge eines Teilnehmenden“. Informationen über die durchschnittlichen Bewertungen eines Teilnehmenden als Indikator für den „Strengegrad bei der Benotung“ eines Teilnehmenden werden nicht berichtet. Bei einer relativen Betrachtung der einzelnen Stores innerhalb eines Unternehmens sollten die Mittelwerte der Bewertungen der Teilnehmenden ähnlich sein. Ohne eine entsprechende Ausweisung ist eine Bewertung dieses Aspekts nicht möglich.

Aufgrund der o. g. Aufführungen ist bei der Messung der Performance von Shoppingcentern von einer geringen Reliabilität und Validität auszugehen. Dies schränkt die Aussagekraft der Befragung deutlich ein. Um eine gesicherte Aussagekraft zu gewährleisten, ist eine höhere Messqualität erforderlich, die auf einer umfassenderen Erhebung der Performance basiert und sorgfältig entwickelte Items verwendet, die von allen Teilnehmenden gleich interpretiert werden. Hinsichtlich der Methodik zur Erhebung von Performance hat es also im SCPR 2023 im Vergleich zu den Vorjahren keine Verbesserung gegeben.

DIE AUSWAHL DER STICHPROBE

Nach Eigenangabe erfolgt die Rekrutierung der Teilnehmenden durch die Studienleiter der ecostra GmbH über die unternehmensinterne Kontaktdatenbank mit E-Mail-Adressen von Vertretern von Filialisten, die in deutschen Shoppingcentern eingemietet sind. Die Datenbank besteht seit 2011. Unterstützend rufen die Medienpartner Immobilien Zeitung und TextilWirtschaft über Print, Newsletter sowie Social Media jährlich zur Teilnahme an der Befragung auf. Dafür ist die Aufnahme in die Kontaktdatenbank der ecostra GmbH erforderlich (vgl. S. 1 f. ecostra 2023).

›Trotz der Bemühungen der ecostra GmbH um eine umfassende Datenbasis ist eine systematische Stichprobenverzerrung bei der Rekrutierung der Befragungsteilnehmer nicht auszuschließen.‹

Trotz der Bemühungen der ecostra GmbH um eine umfassende Datenbasis ist eine systematische Stichprobenverzerrung bei der Rekrutierung der Befragungsteilnehmer nicht auszuschließen. So könnten z. B. die aktive Leserschaft der beiden Medienpartner sowie die Geschäftskontakte der ecostra GmbH überrepräsentiert, kleinere Filialisten aufgrund der schlechteren Erreichbarkeit ggf. unterrepräsentiert sein. Diese und weitere mögliche Stichprobenverzerrungen können jedoch ohne genauere Kenntnis der Datenbasis sowie der Grundgesamtheit möglicher Ansprechpartner/Unternehmen nicht näher bestimmt werden.

Ein weiteres methodisches Problem sind die über die Jahre hinweg gleichen Ansprechpartner der Datenbank. Die Datenbank wird bei personellen Wechseln zwar fortlaufend durch die ecostra GmbH aktualisiert, dennoch sind die Bewertungen der Stores in den Shoppingcentern über die Jahre hinweg nicht unabhängig. Das kann zu einer Überschätzung der vermeintlichen inhaltlichen Konsistenz in den Bewertungen führen.

In die Auswertung gingen 94 vollständige Fragebögen von Teilnehmenden ein (S. 1 in ecostra 2023), die insgesamt mehr als 3.415 Standorte bewerteten. Dies ergibt einen Durchschnitt von 36 Bewertungen pro teilnehmender Person. Für viele Teilnehmende dürfte die Befragung daher sehr zeitaufwändig und ermüdend gewesen sein, was in der Regel die Antwortqualität mindert und die Abbruchwahrscheinlichkeit deutlich erhöht. Dies birgt die Gefahr einer verzerrten Repräsentativität der Stichprobe. Auch dieser Aspekt ist ohne genaue Angaben zu den Abbrechern nicht eindeutig bestimmbar.

DIE STICHPROBENGRÖSSE PRO SHOPPINGCENTER

Die Anzahl der Bewertungen pro Shoppingcenter ist für die Mehrzahl der 400 Center deutlich zu gering für eine aussagekräftige Statistik, oft sogar im einstelligen Bereich (vgl. Tab. 2 ab S. 16 in ecostra 2023). Die durchschnittliche Anzahl der Bewertungen pro Center liegt unter neun. Als Faustformel für die Stichprobengröße werden in der wissenschaftlichen Literatur häufig ca. 30 Datenpunkte (d.h. Personen, Bewertungen etc.) pro Untergruppe genannt (vgl. Kreis et al. 2021, S. 213; Koschate-Fischer und Schandelmeier 2014).¹

1) Hierbei handelt es sich lediglich um eine sehr grobe Faustregel. Die Berechnung einer empfehlenswerten Stichprobengröße hängt von einer Vielzahl an Faktoren sowie Annahmen ab, die je nach Einzelfall unterschiedlich sind. Jedoch ist den Gutachtern keine wissenschaftliche Studie zu Befragungsergebnissen bekannt, bei der eine Stichprobengröße im einstelligen Bereich als ausreichend für belastbare Aussagen angesehen wurde.



Diese Anzahl wird in fast allen Fällen selbst bei großen Shoppingcentern mit vielen Mietern nicht erreicht.

Aufgrund der begrenzten, zum Teil geringen Anzahl von Mietern pro Center, statistisch ausgedrückt eine geringe Grundgesamtheit, kann durchaus argumentiert werden, dass auch mit einer kleineren Stichprobe verwertbare Ergebnisse erzielt werden können. Die von den Studienleitern definierte Mindestzahl von fünf Bewertungen pro Center ist jedoch unabhängig von der Grundgesamtheit statistisch gesehen in jedem Fall deutlich zu gering, um eine fundierte Aussage über die Leistungsfähigkeit zu treffen.

Weiterhin ist zur Stichprobengröße pro Center kritisch anzumerken, dass die absolute Anzahl an Bewertungen bei sehr unterschiedlich großen Centern (von 10.000 bis 120.000 qm Fläche) kein gutes Maß für die Repräsentativität der Mieter darstellt. Ein sehr großes Center mit vielen Mietern kann noch weniger valide auf Basis einer Anzahl ab fünf Bewertungen beurteilt werden als ein kleines Center mit wenigen Mietern. Methodisch sinnvoller wäre es dementsprechend, auch den Anteil der teilnehmenden Mieter pro Center zu berichten und auf dieser Basis zusätzlich den Mindestanteil für eine Bewertbarkeit festzulegen. Der SCPR 2023 weist in dieser Hinsicht eine positive Veränderung zum SCPR 2018 auf: Im Ranking wird nun pro Center neben der Stichprobengröße („Anzahl an Bewertungen“) auch die Mieteranzahl aufgeführt. Leider fehlt der Quotient, aber es ist nun dennoch indirekt möglich, den Anteil der Mieter zu bestimmen, die an der Befragung teilgenommen haben. Diese Neuerung ist positiv zu werten, ändert jedoch nichts am Problem der zu niedrigen Mindestgrenze von fünf Bewertungen, um in das Ranking aufgenommen zu werden.

Darüber hinaus wird die Varianz der Antworten pro Center auf der Skala 1-5 im Report nicht angegeben. In Folge der fehlenden Varianzangaben pro Center ist bspw. auch die Berechnung von Konfidenzintervallen zur Abschätzung der Genauigkeit einer Performancebewertung nicht möglich.² Auch Testverfahren zur statistischen Signifikanz von Mittelwertunterschieden sind aus dem gleichen Grund nicht durchführbar. Mithilfe weniger Annahmen kann jedoch exemplarisch gezeigt werden, dass sich bei geringen Stichprobengrößen im einstelligen Bereich die Mittelwerte von bspw. 3,0 und 2,5 statistisch nicht mit ausreichender Sicherheit unterscheiden lassen.³ Derartige Ungenauigkeiten in der Messung haben jedoch auf den Rangplatz eines Shoppingcenters einen enormen Einfluss. So liegen bspw. im Report von 2023 zwischen den Bewertungen 2,5

»In Folge der fehlenden Varianzangaben pro Center ist bspw. auch die Berechnung von Konfidenzintervallen zur Abschätzung der Genauigkeit einer Performancebewertung nicht möglich.«

2) Konfidenzintervalle, auch Vertrauensintervalle genannt, geben die Präzision einer Lageschätzung eines Parameters auf Basis einer Stichprobe an. So kann bspw. für ein Shoppingcenter ein Intervall um den Mittelwert (als Performance-Schätzung) der Bewertungen berechnet werden, das den „wahren“ Parameter (in diesem Fall den wahren Performance-Score auf der 5-Punkte-Skala) mit einer vordefinierten Wahrscheinlichkeit überdeckt. Ein Konfidenzintervall pro Center wäre im Fall der im SCPR verwendeten ordinalen Skalierung für Performance ebenfalls kritisierbar, aber dennoch ein hilfreicher Indikator für die Genauigkeit der Performance-Schätzung.

3) Exemplarische Annahmen: Stichprobengröße von jeweils $n=9$ pro Center, Mittelwerte von 2,5 und 3,0 sowie eine Varianz von jeweils 0,75, Signifikanzniveau von 10%. Hierbei handelt es sich um eine eher konservative Schätzung.

(Rang 45) und 3,0 (Rang 116, wurde 16-mal wegen gleichem Mittelwert vergeben) 71 Ränge. Zwischen den Bewertungen 3,0 und 3,5 (einschließlich der Intervallgrenzen) liegen 73 Ränge.

Deswegen liegt auch im SCPR 2023 die Vermutung nahe, dass das Ranking sehr stark von Störvariablen sowie dem reinen Zufall aufgrund üblicher Schwankungen bei kleinen Stichproben-
größen beeinflusst wird. Dies wird auch deutlich in Tabelle 5 des Reports (ab S. 40), in der zu erkennen ist, dass vor allem dort große Differenzen von 0,8 bis 1,0 durchschnittlichen Notenpunkten bei den Centern im Vergleich zum Vorjahr auftreten, wo die Stichprobengröße gering ist.⁴ Eine valide Rangbestimmung für ein Center ist damit aus wissenschaftlich-methodischer Sicht nicht möglich.

DARSTELLUNG UND INTERPRETATION DER ERGEBNISSE

Wie bereits 2018 werden auch beim SCPR 2023 die Durchschnittsbewertungen mit zwei Nachkommastellen angegeben und auf dieser Basis das Ranking erstellt. Dies suggeriert eine Messgenauigkeit, die mit dieser Stichprobe nicht annähernd eingehalten werden kann. Wenig verwunderlich ist es bei der häufig geringen Teilnehmerzahl pro Center, dass viele Mittelwerte für Center identisch sind und damit Ränge geteilt werden. Die Untersuchungsleiter erkennen dieses Problem und räumen dies auch offen an mehreren Stellen im Report ein (z. B. wird im Fazit auf S. 25 in ecostra 2023 darauf hingewiesen). Die Interpretation der Ergebnisse ist insgesamt im Vergleich zum SCPR 2018 umsichtiger geworden und weist deutlicher auf die Probleme der Datenerhebung und Darstellung hin. Dies ist als eine positive Entwicklung zu sehen, auch wenn dadurch die mangelhafte Validität der Erhebung nicht korrigiert werden kann.

Als positive Entwicklung von 2018 zu 2013 ist außerdem die Einführung einer neuen Farbcodierung zu sehen. Diese Kodierung gruppiert alle Center, die sich innerhalb eines Intervalls mit der Notenbreite von 0,5 bewegen (z. B. 2,0 bis 2,49). Dadurch wird den Lesern verdeutlicht, nicht zu viel Wert auf die konkreten Ränge zu legen, sondern in größeren Kategorien zu klassifizieren. Dieser Ansatz ist richtig, auch wenn die Intervallbreiten von 0,5 unserer Ansicht nach noch etwas zu eng gefasst sind, solange die Mindestanzahl an Bewertungen bei fünf verbleibt.⁵

Der SCPR 2023 macht im Vergleich zum SCPR 2018 zudem an vielen Stellen noch stärker deutlich, dass es sich um eine Mieterbefragung handelt. Die Untersuchungsleiter betonen, dass sich das nicht ändern soll und die Sichtweisen anderer Akteure nicht vermischt werden sollen (vgl. S. 1 im Vorwort). Die Fokussierung auf eine Zielgruppe ist im SCPR 2023 nun ausreichend transparent dargestellt. Die Bezeichnung der Studie als „Performance Report“ kann zwar noch immer leicht missverstanden werden, aber der Zusatz „Mieterbefragung 2023“ im Titel verdeutlicht nun transparenter den Fokus des Reports. Dies ist ebenfalls eine positive Entwicklung der letzten Jahre.

FAZIT ZUR METHODIK DES SHOPPINGCENTER PERFORMANCE REPORTS DEUTSCHLAND 2023

Die Validität und Reliabilität des SCRPR 2023 ist aus wissenschaftlich-methodischer Sicht als relativ gering einzustufen. Dies ist in erster Linie auf die Performance-Messung sowie die Stichprobengröße zurückzuführen. Des Weiteren kann nicht ausgeschlossen werden, dass es aufgrund der Stichprobenauswahl zu Verzerrungen hinsichtlich der Repräsentativität kommt.

Der Single-Item Ansatz zur Messung der Performance ist nicht empfehlenswert. Die Uneindeutigkeit der Frage sowie die hohe Komplexität erschweren die intersubjektive Vergleichbarkeit der Bewertungen.

Die größtenteils sehr geringen Stichprobenumfänge pro Shoppingcenter führen, selbst im theoretischen Fall einer qualitativ sehr guten Erhebung der Performance, zu einer unvermeidbaren Streuung der Performance-Bewertungen pro Center. Derartige (Zufalls-)Streuungen haben einen großen Einfluss auf das Ranking, so dass eine Rangreihung sowie eine Angabe von Mittelwerten mit zwei Nachkommastellen statistisch nicht sinnvoll erscheinen.

»Wie bereits 2018 werden auch beim SCPR 2023 die Durchschnittsbewertungen mit zwei Nachkommastellen angegeben und auf dieser Basis das Ranking erstellt. Dies suggeriert eine Messgenauigkeit, die mit dieser Stichprobe nicht annähernd eingehalten werden kann.«

»Die Validität und Reliabilität des SCRPR 2023 ist aus wissenschaftlich-methodischer Sicht als relativ gering einzustufen. Dies ist in erster Linie auf die Performance-Messung sowie die Stichprobengröße zurückzuführen.«

4) Exemplarisch in aufsteigender Reihenfolge vom Rangplatz 2023: Göttingen – Kauf Park Göttingen (n = 6, TN-Anteil 12%); Dinslaken – Neutor Galerie (n = 7, TN-Anteil 18%); Friedrichsdorf – Taunus Carré (n = 5, TN-Anteil 19%).

5) Die Kategorisierung mit unterschiedlichen Farben ohne übergangslosen Farbverlauf birgt das Problem, dass Center mit ähnlicher Durchschnittsbewertung in zwei unterschiedliche Kategorien fallen können. Dies könnte den Eindruck größerer Unterschiede zwischen diesen Centern erwecken. Auf dieses Problem wird im Vorwort hingewiesen (vgl. S. 2 der Vorbemerkungen in ecostra 2023).



Positiv anzumerken ist im Vergleich zum SCPR 2018 die umfassendere Aufklärung zur Methodik und den Einschränkungen der Studie, die vorsichtigere Interpretation der Ergebnisse sowie die Kategorisierung in Gruppen durch eine Farbkodierung.

Die insgesamt leicht verbesserte Gesamtqualität des SCPR 2023 sollte jedoch nicht darüber hinwegtäuschen, dass das Ranking und die Mittelwerte der Center wissenschaftlich nicht als valide zu bewerten sind und aus Sicht der Gutachter nicht als Grundlage für weitreichende (De-)Investitionsentscheidungen Anwendung finden sollten.

Kiel, den 19. Februar 2024



© Foto: Hanna Börm

Prof. Dr. Björn Christensen
 (Professor für Statistik und Mathematik)
 Fachhochschule Kiel,
 Sokratesplatz 2, 24149 Kiel



Prof. Dr. Manuel Stegemann
 (Professor für Markt- und Werbepsychologie
 sowie Marketing)
 Hochschule Bielefeld,
 Interaktion 1, 33619 Bielefeld

Quellen

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., Kaiser, S. (2012): Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, 40 (3), 434-449.

ecostra (2018): Shoppingcenter Performance Report Deutschland 2018. Erstellt und vertrieben durch die ecostra GmbH.

ecostra (2023): Shoppingcenter Performance Report Deutschland 2023. Erstellt und vertrieben durch die ecostra GmbH.

Koschate-Fischer, N., Schandelmeier, S. (2014): A guideline for designing experimental studies in marketing research and a critical discussion of selected problem areas. *Journal of Business Economics*, 84 (6), 793-826.

Kreis, H., Wildner, R., Kuß, A. (2021): Marktforschung. Datenerhebung und Datenanalyse, 7. Auflage. Wiesbaden: Springer Gabler.

Moosbrugger, H., Kelava, A. (2012): Testtheorie und Fragebogenkonstruktion, 2. Auflage. Berlin: Springer.

Sarstedt, M., Wilczynski, P. (2009): More for Less? A Comparison of Single-Item and Multi-Item Measures. *Die Betriebswirtschaft*, 69 (2), 211-227.

Kurzvita
Prof. Dr. Björn Christensen

Seit 2020: Präsident der Fachhochschule Kiel
2018 – 2020: Dekan des Fachbereichs Wirtschaft der Fachhochschule Kiel
2016 – 2018: Prodekan für Forschung und Lehre am Fachbereich Wirtschaft der Fachhochschule Kiel
Seit 2013: Professor für Statistik und Mathematik am Fachbereich Wirtschaft der Fachhochschule Kiel
2009 – 2013: Gründer und Geschäftsführer der meteolytix GmbH
2005 – 2013: Gründer und Geschäftsführer der analytix GmbH
2004: Promotion zum Dr. sc. pol. bei Prof. Dr. h.c. Horst Siebert
2000 – 2005: Wissenschaftlicher Mitarbeiter am Kieler Institut für Weltwirtschaft, Schwerpunkte im Bereich der statistischen Analyse von Individualdaten
1995 – 2000: Studium der Volkswirtschaft mit quantitativem Schwerpunkt an der CAU zu Kiel, Abschluss Diplom

Kurzvita
Prof. Dr. Manuel Stegemann

Ab 2021: Professor für Markt- und Werbepsychologie sowie Marketing an der Hochschule Bielefeld
2018 – 2021: Professor für Marketing und Statistik am Fachbereich Wirtschaft der Fachhochschule Kiel
Seit 2018: Professor für Marketing und Statistik am Fachbereich Wirtschaft der Fachhochschule Kiel
2014 – 2018: Unternehmensberater und Partner bei der Beratung Kappes & Partner, Schwerpunkte: HR und Marketing im Gesundheitswesen
2011 – 2014: Wissenschaftlicher Mitarbeiter beim Institut für Marketing der Universität Münster (Prof. Dr. M. Krafft), Kumulative Promotion zum Thema: *Success Factors of Pay What You Want Pricing*
2010 – 2011: Unternehmensberater bei der Boston Consulting Group, Schwerpunkte Marketing & Sales, HR
2004 – 2009: Studium der Psychologie an der Westfälischen Wilhelms-Universität Münster, Schwerpunkte: Statistik und Methodenlehre, Arbeits- und Organisationspsychologie

RÜCKBLICK

GUTACHTEN ZUR METHODIK DES SHOPPINGCENTER PERFORMANCE REPORT DEUTSCHLAND 2018 (ECOSTRA GMBH)

Prof. Dr. Björn Christensen und **Prof. Dr. Manuel Stegemann**
(Fachhochschule Kiel) im Auftrag des German Council of
Shopping Centers e. V.



Gutachten zur Methodik des Shoppingcenter Performance Report Deutschland 2018

Die Methodik des Shoppingcenter Performance Reports (SCPR) des Beratungsunternehmens ecostra GmbH 2018 soll im Rahmen dieses Gutachtens hinsichtlich der zugrundeliegenden Methodik wissenschaftlich bewertet werden. Dies erfolgt insbesondere hinsichtlich des Anspruchs seitens der Herausgeber, dass der SCPR ein valides Ranking der wirtschaftlichen Performance von 400 deutschen Shoppingcentern aus Mietersicht abbildet (vgl. S. 2 der Vorbemerkungen des SCPR 2018).

KURZBESCHREIBUNG DES SHOPPINGCENTER PERFORMANCE REPORTS DEUTSCHLAND 2018

Der Report, dessen Hauptbestandteil ein Ranking von 400 Shoppingcentern in Deutschland bildet, wird seit dem Jahr 2011 jährlich von der ecostra GmbH, einer Wirtschafts-, Standort- und Strategieberatung, erstellt und kostenpflichtig vertrieben.

Als Ziel des Reports wird seitens der ecostra GmbH genannt, dass den Marktteilnehmern in einer systematischen und Überblickshafter Weise Informationen und Nutzungsmöglichkeiten bereitgestellt werden bzgl. der wirtschaftlichen Performance von Shoppingcentern aus Mietersicht (vgl. S. 2 der Vorbemerkungen des Shoppingcenter Performance Reports Deutschland 2018). Die Datengrundlage für den Report bildet eine Online-Umfrage unter Ansprechpartnern von Unternehmen, die als Mieter Stores in deutschen Shoppingcentern betreiben. Die Umfrage wurde von ecostra im Zeitraum vom 28.06.2018 bis zum 02.09.2018 durchgeführt.

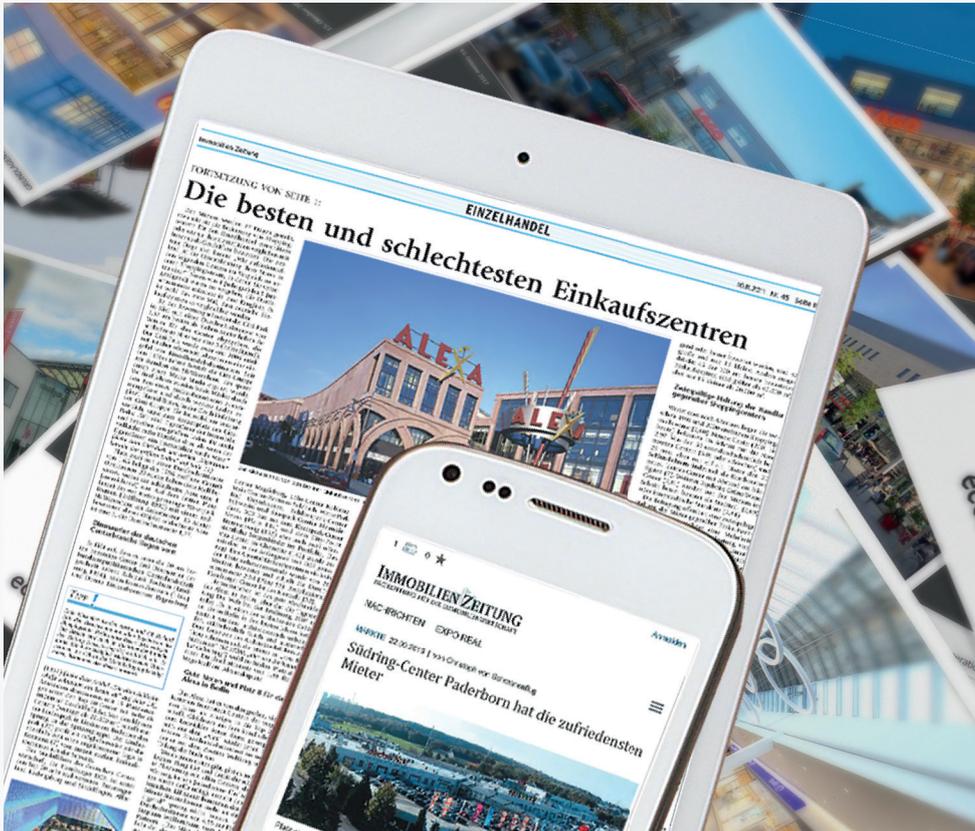
METHODIK ZUR ERHEBUNG DER PERFORMANCE VON SHOPPINGCENTERN

Im vorliegenden Fall des SCPR wurde die Performance eines Stores in einem Shoppingcenter anhand einer einzigen Frage gemessen (S. 14 im SCPR 2018). In der wissenschaftlichen Literatur wird ein sog. Single-Item-Measurement nur in Ausnahmefällen befürwortet, wenn das zu messende Merkmal sehr konkret und homogen, d. h. wenig facettenreich, ist (vgl. Kuß et al. 2018, S. 99ff.; Diamantopoulos et al. 2012). Die Performance eines Shoppingcenters ist jedoch zweifellos ein sehr heterogenes, komplexes Merkmal mit verschiedenen Facetten. Es ist daher auf Basis wissenschaftlicher Qualitätskriterien nicht empfehlenswert, dieses Merkmal mit nur einer einzigen Frage zu messen. Zahlreiche Studien zeigen, dass in einem derartigen Fall sowohl die Reliabilität (Zuverlässigkeit) als auch die Validität (Gültigkeit) einer Messung eingeschränkt sind und damit Ergebnisse nur mit großer Vorsicht zu interpretieren sind (vgl. z.B. Diamantopoulos et al. 2012; Sarstedt, M. und Wilczynski, P. 2009).

Es ist demzufolge von einer relativ geringen Reliabilität und Validität bei der Messung der Performance im Falle von Shoppingcentern auszugehen. Dies schränkt die Aussagekraft der Befragung deutlich ein. Für eine gesicherte Aussagekraft der Befragung ist eine höhere Messqualität und damit eine umfassendere Erhebung der Performance auf Basis mehrerer, sorgfältig entwickelter Items angebracht.

Die verwendete Frage zur Messung der Performance ist zudem kompliziert formuliert und erfordert eine komplexe Bewertung verschiedener Faktoren seitens der Teilnehmenden: So soll die Umsatzleistung eines einzelnen Stores in Relation zu den Standortkosten und gleichzeitig relativ zu anderen Stores des Unternehmens in Shoppingcentern gesetzt werden. Eine derart komplexe Frage, die verschiedene Faktoren verknüpft, dürfte von Teilnehmenden unterschiedlich verstanden werden, was wiederum zu intersubjektiv verschiedenen Bewertungsdimensionen der Teilnehmenden führt.

›Zahlreiche Studien zeigen, dass in einem derartigen Fall sowohl die Reliabilität (Zuverlässigkeit) als auch die Validität (Gültigkeit) einer Messung eingeschränkt sind und damit Ergebnisse nur mit großer Vorsicht zu interpretieren sind.‹



Die Eindeutigkeit einer Fragestellung ist jedoch erforderlich, um eine intersubjektiv gemeinsame Verständnisbasis zu schaffen und um die Ergebnisse vergleichbar zu machen (vgl. z.B. Moosbrugger, H. und Kelava, A. 2012, S. 65f.).

Weiterhin beinhaltet die Frage einen relativen Vergleich zu den anderen Stores des jeweiligen Unternehmens. Dies ist methodisch als problematisch anzusehen, da jedes teilnehmende Unternehmen eine andere Vergleichsbasis hat, insbesondere hinsichtlich der individuellen Standortkosten, was die Vergleichbarkeit der Bewertungen erheblich schmälert. Für den Report ist jedoch eine hohe Vergleichbarkeit besonders wichtig, da die Shoppingcenter in eine Rangfolge auf Basis der Durchschnittsbewertung mit zwei Nachkommastellen gebracht werden.

Die verwendete Skala zur Bewertung ist angelehnt an ein Schulnotensystem (1 = sehr gut bis 5 = mangelhaft), der Inhalt der Frage bezieht sich jedoch auf die monetäre Situation der einzelnen Stores relativ zu den unternehmensinternen vergleichbaren Stores in Shoppingcentern. Eine Überführung in ein Schulnotensystem ist zwar nicht unüblich in Befragungen, jedoch unterliegt diese Art der Bewertung dem Einfluss der „individuellen Strenge eines Teilnehmenden“. Informationen zu den durchschnittlichen Bewertungen eines Teilnehmenden als Indikator für den „Strengegrad bei der Benotung“ eines Teilnehmenden werden nicht berichtet. Bei einer relativen Betrachtung der einzelnen Stores innerhalb des Unternehmens sollten sich die Mittelwerte der Bewertungen der Teilnehmenden ähneln. Ohne eine entsprechende Ausweisung ist eine Bewertung dieses Aspekts nicht möglich.

In Bezug auf die Frageformulierung und -komplexität muss insgesamt aufgrund der o. g. Ausführungen von einer geringen Reliabilität und Validität sowie weiterhin von einer geringen Vergleichbarkeit der Bewertungen ausgegangen werden.

DIE AUSWAHL DER STICHPROBE¹

Die Rekrutierung der Teilnehmenden erfolgt durch die Studienleiter der ecostra GmbH über die unternehmensinterne Kontaktdatenbank mit Email-Adressen von Vertretern von Filialisten, die in deutschen Shoppingcentern eingemietet sind. Die Datenbank

»Bei einer relativen Betrachtung der einzelnen Stores innerhalb des Unternehmens sollten sich die Mittelwerte der Bewertungen der Teilnehmenden ähneln. Ohne eine entsprechende Ausweisung ist eine Bewertung dieses Aspekts nicht möglich.«

besteht seit 2011. Unterstützend wird über die Medien zur Teilnahme an der Befragung aufgerufen. Für die Teilnahme ist die Aufnahme in die Kontaktdatenbank der ecostra GmbH erforderlich.

Trotz Bemühungen der ecostra GmbH zur Erstellung einer umfassenden Datenbank ist in diesem Falle eine systematische Stichprobenverzerrung bei der Rekrutierung von Umfrageteilnehmern nicht ausgeschlossen. Die aktive Leserschaft der beiden Medienpartner sowie Geschäftskontakte der ecostra GmbH könnten bspw. überrepräsentiert sein, kleinere Filialisten aufgrund der schwereren Erreichbarkeit ggf. unterrepräsentiert. Diese sowie weitere mögliche Stichprobenverzerrungen sind jedoch ohne nähere Kenntnis der Datenbank sowie der Population an möglichen Ansprechpartnern/Unternehmen nicht näher bestimmbar.

Ein weiteres methodisches Problem sind die über die Jahre hinweg gleichen Ansprechpartner der Datenbank. Die Datenbank wird bei personellen Wechseln zwar fortlaufend durch die ecostra GmbH aktualisiert, dennoch sind die Bewertungen der Stores in den Shoppingcentern über die Jahre hinweg nicht unabhängig. Das kann zu einer Überschätzung der vermeintlichen inhaltlichen Konsistenz in den Bewertungen führen, die möglicherweise eher auf die persönliche Meinung der Antwortenden zu den Stores (und damit auch zu den Centern) als auf objektive Einschätzungen zurückzuführen sein dürfte.

Es wurden für den SCPR 2018 rund 800 Ansprechpartner per Email angeschrieben, davon sind nach Angaben der ecostra GmbH etwa 600 bis 650 Email-Adressen gültig. Es wurden 95 vollständige Fragebögen von Teilnehmenden in die Analyse einbezogen (S. 1 des SCPR 2018), die insgesamt 3.653 Store-Standorte bewertet haben. Das ergibt durchschnittlich 38,5 Store-Bewertungen pro Teilnehmenden. Der Anteil abgebrochener Befragungen liegt nach Schätzungen der ecostra GmbH bei etwa 20%. Es ist nach Rücksprache mit der ecostra GmbH aufgrund der vielen erforderlichen Store-Bewertungen durchaus möglich, dass Filialisten mit vielen Stores die Umfrage überproportional häufig abbrechen oder nicht alle Stores in Shoppingcentern bewerten. Dies könnte die Repräsentativität der Stichprobe wiederum verzerren.

Der Effekt durch den Ausschluss von Teilnehmenden aufgrund von weniger als drei Stores in deutschen Shoppingcentern ist mit maximal fünf Fällen pro Jahr recht begrenzt. Auch wird seitens der ecostra GmbH auf doppelte Ansprechpartner pro Unternehmen geachtet, sodass diesbezüglich von keiner weiteren bedeutenden Verzerrung auf die Stichprobe ausgegangen wird.

DIE STICHPROBENGRÖSSE PRO SHOPPINGCENTER

»Die Anzahl an Bewertungen pro Shoppingcenter ist für die Mehrheit der 400 Center deutlich zu gering für eine aussagekräftige Statistik«

Die Anzahl an Bewertungen pro Shoppingcenter ist für die Mehrheit der 400 Center deutlich zu gering für eine aussagekräftige Statistik, sie liegt oft sogar im einstelligen Bereich (ab S. 15 im SCPR 2018). Die durchschnittliche Anzahl an Bewertungen pro Center liegt bei neun. Als gängige Faustregel für die Stichprobengröße werden in der wissenschaftlichen Literatur oft etwa 30 Datenpunkte (d.h. Personen, Bewertungen etc.) pro Untergruppe genannt (vgl. Kuß et al. 2018, S. 204; Koschate-Fischer und Schandelmeier 2014).² Diese Anzahl wird in nahezu allen Fällen selbst bei großen Shoppingcentern mit vielen Mietern nicht erreicht. Aufgrund einer limitierten Anzahl von Mietern pro Center, statistisch ausgedrückt aufgrund einer geringen Populationsgröße, kann durchaus argumentiert werden, dass mit einer kleineren Stichprobe ebenfalls verwendbare Ergebnisse erzielt werden können. Jedoch ist die von den Studienleitern definierte Mindestanzahl von fünf Bewertungen pro Center aus statistischer Sicht unabhängig von der Populationsgröße in jedem Fall deutlich zu gering für eine fundierte Aussage zur Leistungsfähigkeit.

Weiterhin ist zur Stichprobengröße pro Center kritisch anzumerken, dass die absolute Anzahl an Bewertungen bei sehr unterschiedlich großen Centern (von 10.000 bis 120.000 qm Fläche) kein gutes Maß für die Repräsentativität der Mieter darstellt. Ein sehr großes Center mit vielen Mietern kann noch weniger valide auf Basis einer Anzahl ab fünf Bewertungen beurteilt werden als ein kleines Center. Methodisch sinnvoller wäre es dementsprechend, auch den Anteil der teilnehmenden Mieter pro Center zu

berichten und auf dieser Basis zusätzlich den Mindestanteil für eine Bewertbarkeit festzulegen. Eine zweite Alternative dazu wäre, den Anteil der Verkaufsfläche der an der Umfrage teilnehmenden Mieter zu bestimmen und als Kriterium für die Bewertbarkeit eines Centers heranzuziehen. Es sei aber angemerkt, dass beide Vorschläge nichts an dem Problem einer möglichen systematischen Stichprobenverzerrung ändern.

Darüber hinaus wird die Varianz der Antworten pro Center (Skala 1 bis 5) im Report nicht angegeben, nur die im Allgemeinen recht wenig aussagekräftige Spannweite auf der verwendeten 5er-Skala wird pro Center berichtet (ab S. 15 im SCPR 2018). Die Spannweite berücksichtigt – im Gegensatz zur Varianz – nur die zwei extremsten Bewertungen und lässt alle anderen Bewertungen unberücksichtigt. Daher ist sie kein empfehlenswertes Maß für die Darstellung der Streuung von Bewertungen.

In Folge der fehlenden Varianzangaben pro Center ist bspw. auch die Berechnung von Konfidenzintervallen zur Abschätzung der Genauigkeit einer Performancebewertung nicht möglich.³ Auch Testverfahren zur statistischen Signifikanz von Mittelwertunterschieden sind aus dem gleichen Grund nicht durchführbar. Mithilfe weniger Annahmen kann jedoch exemplarisch gezeigt werden, dass sich bei geringen Stichprobengrößen im einstelligen Bereich die Mittelwerte von bspw. 3,0 und 2,5 statistisch nicht mit ausreichender Sicherheit unterscheiden lassen.⁴ Derartige Ungenauigkeiten in der Messung haben jedoch auf den Rangplatz eines Shoppingcenters im Ranking des Reports 2018 einen enormen Einfluss. So liegen bspw. 98 Ränge zwischen den Bewertungen 2,5 und 3,0 (einschließlich der Intervallgrenzen) sowie 56 Ränge zwischen den Bewertungen 3,0 und 3,5 (einschließlich der Intervallgrenzen).

Es liegt dementsprechend die Vermutung nahe, dass das Ranking sehr stark von Störvariablen sowie dem reinen Zufall aufgrund üblicher Schwankungen bei kleinen Stichprobengrößen beeinflusst wird. Dies wird auch deutlich in Tabelle 5 des Reports (ab S. 37), in der zu erkennen ist, dass vor allem dort große Differenzen von 0,5 bis 1,0 Notenpunkten bei den Centern im Vergleich zum Vorjahr auftreten, wo die Stichprobengröße gering ist.⁵ Eine valide Rangbestimmung für ein Center ist damit aus wissenschaftlich-methodischer Sicht nicht möglich.

Die Studienleiter scheinen das Problem teils selbst erkannt zu haben: Es wird im Report (S. 26 im SCPR 2018) von einer sehr großen Spreizung der Performance-Bewertungen, sowohl bei sehr gut platzierten also auch bei am unteren Ende platzierten Centern, berichtet. Diese Spreizung, die abgesehen von den berichteten Spannweiten nicht näher beschrieben wird, ist ein Indikator für die Ungenauigkeit der Messmethodik und somit auch für die Ungenauigkeit des erstellten Rankings. Die Studienleiter räumen korrekterweise ebenfalls ein, dass sich an vielen Stellen des Rankings sehr viele Ränge auf nur eine geringe Differenz der Durchschnittsbewertungen verteilen (S. 24 des SCPR 2018).

Weiterhin suggeriert das Berichten von zwei Nachkommastellen des Mittelwertes für die Aufstellung der Rangfolge eine Messgenauigkeit, die mit dieser Stichprobe nicht eingehalten werden kann. Es wäre empfehlenswert, ein Center erst mit einer deutlich

›In Folge der fehlenden Varianzangaben pro Center ist bspw. auch die Berechnung von Konfidenzintervallen zur Abschätzung der Genauigkeit einer Performancebewertung nicht möglich. Auch Testverfahren zur statistischen Signifikanz von Mittelwertunterschieden sind aus dem gleichen Grund nicht durchführbar.‹

¹Die im Folgenden berichteten Angaben zur Stichprobenauswahl sind größtenteils nicht im SCPR 2018 ersichtlich, sondern wurden freundlicherweise von der ecostra GmbH auf Nachfrage per Email am 28. Oktober 2019 bereitgestellt.

²Hierbei handelt es sich lediglich um eine sehr grobe Faustregel. Die Berechnung einer empfehlenswerten Stichprobengröße hängt von einer Vielzahl an Faktoren sowie Annahmen ab, die je nach Einzelfall unterschiedlich sind. Jedoch ist den Gutachtern keine wissenschaftliche Studie zu Befragungsergebnissen bekannt, bei der eine Stichprobengröße im einstelligen Bereich als ausreichend für belastbare Aussagen angesehen wurde.

³Konfidenzintervalle, auch Vertrauensintervalle genannt, geben die Präzision einer Lageschätzung eines Parameters auf Basis einer Stichprobe an. So kann bspw. für ein Shoppingcenter ein Intervall um den Mittelwert (als Performance-Schätzung) der Bewertungen berechnet werden, das den „wahren“ Parameter (in diesem Fall den wahren Performance-Score) mit einer vordefinierten Wahrscheinlichkeit überdeckt. Ein Konfidenzintervall pro Center wäre im Fall der im SCPR verwendeten ordinalen Skalierung für Performance ebenfalls kritisierbar, aber dennoch ein hilfreicher Indikator für die Genauigkeit der Performance-Schätzung.

⁴Exemplarische Annahmen: Stichprobengröße von jeweils $n=9$ pro Center, Mittelwerte von 2,5 und 3,0 sowie eine Varianz von jeweils 0,75, Signifikanzniveau von 10%. Hierbei handelt es sich um eine eher konservative Schätzung.

größeren Anzahl an Bewertungen in eine Vergleichsanalyse aufzunehmen und weiterhin anstatt eines Rankings eine gröbere Klassifizierung zur Performance der Center vorzunehmen. Dies würde der Ungenauigkeit des derzeitigen Messverfahrens gerecht werden und die Aussagekraft erhöhen.

DIE INTERPRETATION DER ERGEBNISSE

Es wird seitens der Studienleiter an verschiedenen Stellen im Bericht (z. B. ab S. 24 des SCPR 2018) betont, dass die Ergebnisse in verschiedener Hinsicht eine hohe Konstanz zu den Vorjahren aufweisen. Diese kann – wie im Abschnitt zur Auswahl der Stichprobe bereits erläutert – bei genauerer Betrachtung auch eine Folge der Methodik und weniger der inhaltlichen Konsistenz der Messung sein: Da die Befragung in den Jahren überwiegend von den gleichen Ansprechpartnern der Unternehmen beantwortet wurde, ist eine Konsistenz im Ranking möglicherweise auch auf die persönliche Meinung der Antwortenden zurückzuführen und weniger auf eine inhaltlich valide Messung. Unterschiede zwischen den bewerteten Stores können möglicherweise prägnant auf interstorespezifische Unterschiede zwischen den Ansprechpartnern je Store in der Auffassung über die Fragestellung (vgl. ersten Aspekt) mit mehreren Dimensionen erklärt werden und weniger mit objektiven Unterschieden zwischen den Stores. Eine die Methodik und das Reporting betreffende Fragestellung ist zudem, inwiefern die im Ranking aufgeführten Center vergleichbar sind. Das Hauptaugenmerk wird auf das nicht kategorisierte Ranking aller Center gelegt, das zu Beginn des Reports (ab S. 15 im SCPR 2018) dargestellt wird und auf dessen Basis die Hauptschlussfolgerungen gezogen werden. Die Sonderauswertung für Bahnhöfe und Flughäfen (ab S. 35 des SCPR 2018) erscheint sinnvoll, ebenso wie die Gruppierung der Center nach Geschäftsfläche in drei Klassen (dargestellt in den Tabellen 6 bis 8 ab S. 49). Jedoch bleibt der Eindruck, dass eine feinere Kategorisierung nach thematischen Schwerpunkten und der Lage (z. B. ländlich, städtisch wegen der unterschiedlichen Standortfaktoren) zwecks einer besseren Vergleichbarkeit empfehlenswert ist. Die wesentlichen Schlussfolgerungen können dann aus dem Vergleich ähnlicherer Center gewonnen werden.⁶

›Da die Befragung in den Jahren überwiegend von den gleichen Ansprechpartnern der Unternehmen beantwortet wurde, ist eine Konsistenz im Ranking möglicherweise auch auf die persönliche Meinung der Antwortenden zurückzuführen und weniger auf eine inhaltlich valide Messung.‹

›Letztlich trifft die Bezeichnung der Studie als Performance Report inhaltlich nur sehr bedingt zu. Anknüpfend an die bereits dargestellte Problematik der Erfassung der Performance mit nur einer einzigen Frage spiegelt sie auch lediglich die Sichtweise der Mieter wider.‹

Letztlich trifft die Bezeichnung der Studie als Performance Report inhaltlich nur sehr bedingt zu. Anknüpfend an die bereits dargestellte Problematik der Erfassung der Performance mit nur einer einzigen Frage spiegelt sie auch lediglich die Sichtweise der Mieter wider. Insofern kann der Titel der Studie als Performance Report leicht missverstanden werden. Die Studie entspricht eher dem Charakter einer Mieterzufriedenheitsbefragung. Es sei aber angemerkt, dass die Studienleiter diese Problematik aktiv aufgreifen und an verschiedenen Stellen (erstmalig auf S. 1 der Vorbemerkung des SCPR 2018) deutlich machen, dass es sich um eine wirtschaftliche Bewertung aus Mietersicht handelt.

FAZIT ZUR METHODIK DES SHOPPINGCENTER PERFORMANCE REPORTS DEUTSCHLAND 2018

Die Validität und Reliabilität des SCRP 2018 ist aus wissenschaftlich-methodischer Sicht als relativ gering einzuschätzen. Dies liegt in erster Linie an der Messung der Performance sowie der Stichprobengröße. Nachrangig kann auch nicht ausgeschlossen werden, dass durch die Stichprobenauswahl Verzerrungen in Bezug auf die Repräsentativität vorliegen.

Der Single-Item-Ansatz zur Messung der Performance ist nicht empfehlenswert. Die Uneindeutigkeit der Frage sowie die hohe Komplexität machen die Bewertungen intersubjektiv schwer vergleichbar.

⁵ In aufsteigender Reihenfolge vom Rangplatz 2018: Chemnitz – Neefepark (n=5), Nordhausen – Südharz Galerie (n=5), Baden-Baden – Shopping Cité (n=6), Stuttgart – Carré Bad Cannstatt (n=5), Berlin – Allee-Center (n=8), Krefeld – Schwanenmarkt (n=9), etc.

⁶ Die Bestimmung der Kriterien zur Kategorisierung vergleichbarer Shoppingcenter bedarf fachliches Expertenwissen und wird dementsprechend in diesem Gutachten nicht tiefergehend thematisiert.

Die größtenteils sehr geringen Stichprobengrößen pro Shoppingcenter sorgen, selbst im Falle einer qualitativ sehr guten Erhebung der Performance, für eine unvermeidliche Streuung der Performance-Bewertungen pro Center. Derartige (Zufalls-)Streuungen üben einen großen Einfluss auf das Ranking aus, sodass eine Rangreihung statistisch nicht sinnvoll erscheint.

Kiel, 31. Oktober 2019 – im Auftrag des German Council of Shopping Centers e. V.



Prof. Dr. Björn Christensen,
Professor für Statistik und Mathematik,
Fachhochschule Kiel



Prof. Dr. Manuel Stegemann,
Professor für Marketing und Statistik,
Fachhochschule Kiel

QUELLEN

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., Kaiser, S. (2012): Guidelines for choosing between multiitem and singleitem scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, 40 (3), 434-449.

Ecostra (2018): Shoppingcenter Performance Report Deutschland 2018. Erstellt und vertrieben durch die ecostra GmbH.

Koschate-Fischer, N., Schandemeier, S. (2014): A guideline for designing experimental studies in marketing research and a critical discussion of selected problem areas. *Journal of Business Economics*, 84 (6), 793-826.

Kuß, A., Wildner, R., Kreis, H. (2018): Marktforschung, Datenerhebung und Datenanalyse, 6. Auflage. Wiesbaden: Springer Gabler.

Moosbrugger, H., Kelava, A. (2012): Testtheorie und Fragebogenkonstruktion, 2. Auflage. Berlin: Springer.

Sarstedt, M., Wilczynski, P. (2009): More for Less? A Comparison of Singleitem and Multiitem Measures. *Die Betriebswirtschaft*, 69 (2), 211-227.

Drei Fragen an:

Ralf-Peter Koschny zu den Ergebnissen des Shoppingcenter Performance Reports

Ralf-Peter Koschny, Sprecher des Vorstandes der bulwiengesa AG und dort verantwortlich für die Bereiche Einzelhandel und Freizeit, bezüglich seiner persönlichen Einschätzung zur Methodik und den Ergebnissen des Shoppingcenter Performance Reports (SCPR).



›Zur Abbildung der tatsächlichen (Umsatz-)Performance eines Centers und damit Qualität eines Centers reicht der SCPR nicht aus.‹

Ralf-Peter Koschny CRE FRICS,
bulwiengesa AG

›Das Ranking hat für Marktteilnehmer, die sich auch nur ansatzweise mit der Methodik auseinandergesetzt haben, keine Aussagekraft.‹

Ralf-Peter Koschny CRE FRICS,
bulwiengesa AG

1. Ein international tätiger Investor möchte ein Shoppingcenter in Deutschland kaufen und beabsichtigt den Shoppingcenter Performance Report auch zu Rate zu ziehen. Welche Entscheidungshilfe bietet der SCPR aus Ihrer Sicht einem Investor?

Internationale Investoren führen beim Kauf eines Shoppingcenters eine commercial und technische Due Diligence durch, die Chancen und Risiken des Objektes analysieren. Der SCPR ist ein Stimmungsbild von wenigen Mietern und wird mit Sicherheit von Investoren, die sich mit der Methode des SCPR auseinandergesetzt haben, nicht ernst genommen. Somit bietet der SCPR keine Entscheidungshilfe. Im Gegenteil, aufgrund des methodisch nicht abgesicherten Rankings dürfte der Report mehr Fragen aufwerfen, als das er Antworten liefert.

Das Stimmungsbild wird nur durch einen Teil der Mieter gebildet. Erfahrungsgemäß antworten die Mieter mit der höchsten Unzufriedenheit und die mit der höchsten Zufriedenheit und die in der Mitte fehlen. Zur Abbildung der tatsächlichen (Umsatz-)Performance eines Centers und damit Qualität eines Centers reicht der SCPR damit nicht aus.

2. Welche Aussagekraft hat das ermittelte Ranking?

Das Ranking hat für Marktteilnehmer, die sich auch nur ansatzweise mit der Methodik auseinandergesetzt haben, keine Aussagekraft. Zur Erklärung: Die Fallzahl ist nicht repräsentativ, das betrifft die gesamte Befragung wie auch die Ergebnisse, die auf einzelne Center heruntergebrochen sind.

3. Welche Stichprobengröße und welche Untersuchungsmethode würde zu wirklich belastbaren Ergebnissen führen?

Wie immer kommt es auf die Art, Größe und die Wettbewerbssituation des jeweiligen Centers an. Wichtig ist, dass hinreichende und qualifizierte Antworten für das jeweilige Center vorliegen. Für uns hätte es eine befriedigende bis gute Aussagekraft, wenn 25 Prozent der Mieter antworten. Darüber hinaus ist zur Einschätzung einer Performance auch die Qualität der Antworten wichtig, d.h. es sollte Wert darauf gelegt werden, dass die bedeutenden Mieter, sprich die Ankermieter, antworten.

Bisher haben sich folgende Methoden positiv für belastbare Ergebnisse einer Performance für ein Shoppingcenter herausgestellt:

- Frequenzählungen im Center (Türzählanlagen) im zeitlichen Ablauf („Längsschnittanalyse“, d.h. über mehrere Jahre verfolgend).
- Evaluation der Miet-Umsatz-Belastung nach Mietern und Branchen: Insbesondere unter Berücksichtigung der durchschnittlichen Flächenproduktivität eines Mieters

und den Vergleich mit seinen Durchschnittswerten (Überdurchschnittlicher/unterdurchschnittlicher Umsatz) in Kombination mit seiner tatsächlich erzielten Effort-Rate (Rent-to-Sales Ratio) und dem Abgleich mit einer centertypischen RSR kann ermittelt werden, wie sich die Anteile im Center hinsichtlich eines Over- oder Underrent darstellen. Dieses Maß ist nach wie vor das Maß aller Dinge zur Ermittlung der Center Performance.

Ralf-Peter Koschny ist Sprecher des Vorstandes der bulwiengesa AG und verantwortlich für die Bereiche Einzelhandel und Freizeit. Er ist seit 1995 in leitender Position bei bulwiengesa. Zuvor war er als Stadtplaner bei FPB – Freie Planungsgruppe Berlin GmbH und als Projektleiter Wirtschaftsförderung der Freien und Hansestadt Hamburg beschäftigt.

Beim GCSC Academy Bord (Forschung/Lehre/Think Tank/Zukunftsfragen) verantwortet er zusammen mit Prof. Dr. Tobias Just (IREBS) den Bereich Research.

Prof. Manfred Güllner: Anmerkungen zum SCPR 2018

Dass die Vorgehensweise bei der Erhebung der Daten des SCPR in keinsten Weise irgendwelchen Regeln der empirischen Forschung entspricht, zeigt schon ein Blick auf die publizierten Ergebnisse des SCPR 2018. Da wird berichtet, dass 30,4 Prozent der Teilnehmer das Centro Oberhausen, 11,6 das Alstertal Einkaufszentrum in Hamburg und 10,1 Prozent Alexa in Berlin als Center mit den höchsten Mieten nennen.

Bei 95 Teilnehmern der „Studie“ entsprechen 10,1 Prozent 9,6 Teilnehmern, 11,6 Prozent 11,0 Teilnehmern. Das zeigt, wie absurd es ist, Prozentanteile hinter dem Komma anzugeben. Es zeigt auch, dass es zwischen 11,6 und 10,1 Prozent eigentlich keinen Unterschied gibt (1 Teilnehmer!) und deshalb auch nicht irgendein Unterschied suggeriert werden darf. Generell gilt dieser Einwand für den gesamten SCPR; denn wenn hier auf der Basis von zum Teil nur 5 Mietern eines Centers zwei Nachkommastellen angegeben und ein Unterschied zwischen einem Wert von 1,56 und einem Wert von 1,57 unterstellt wird, ist das eine krasse Irreführung. In Wirklichkeit kann auf dieser Basis keine Rangfolge erstellt werden.

Bedenken bestehen gegen die Aussagefähigkeit des SCPR auch deshalb, weil in die Bewertung sehr unterschiedliche Bewertungsgrundlagen einfließen. So schwankt die Zahl der Teilnehmer 2018 pro Center zwischen 5 und 30, 2017 gar zwischen 5 und 35. Während also bei einer Basis von nur 5 Teilnehmern ein Teilnehmer das Ergebnis zu 20 Prozent beeinflussen kann, hat bei 30 Teilnehmern ein Teilnehmer nur einen „statistischen Wert“ von 3 Prozent.

Generell ist eine Zahl von 5 Mietern pro Center und eine Teilnehmerzahl von 95 für 259 Center eine viel zu schwache Basis, um statistisch abgesicherte Ergebnisse zu erhalten. Die im SCPR ausgewiesenen Ergebnisse für die einzelnen Center können wegen der geschilderten und auch der bereits in anderen kritischen Beiträgen diskutierten gravierenden methodischen Mängel keinerlei belastbare und realitätsnahe Ergebnisse für ein einzelnes Center liefern. Sie sind – hart formuliert – Artefakte oder Fantazahlen, die ecostra mithilfe von undurchsichtigen Rechnereien produziert.



›Die Vorgehensweise bei der Erhebung der Daten des SCPR entspricht in keinsten Weise irgendwelchen Regeln der empirischen Forschung.‹

Professor Manfred Güllner,
Forsa-Institut

**Gutachten zur Methodik des
Shoppingcenter Performance Report Deutschland 2023
des Beratungsunternehmens ecostra GmbH**

Gutachter:

Prof. Dr. Björn Christensen
(Professor für Statistik und Mathematik)
Fachhochschule Kiel, Sokratesplatz 2, 24149 Kiel

Prof. Dr. Manuel Stegemann
(Professor für Markt- und Werbepsychologie sowie Marketing)
Hochschule Bielefeld, Interaktion 1, 33619 Bielefeld

Datum des Gutachtens:

19. Februar 2024

Auftraggeber:

German Council of Shopping Places e.V.
Mehringdamm 48, D-10961 Berlin
office@gcsp.de