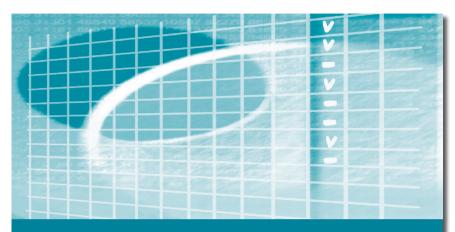
DIGITALER SONDERDRUCK

Zeitschrift für Evaluation, 21. Jahrgang, Heft 1/2022 ISSN 1619-5515

> © Waxmann Verlag GmbH, 2022



Zeitschrift für Evaluation

• Theorie, Methoden und Praxis der Evaluation

Evaluation von Lehrveranstaltungen im Hinblick auf das induzierte эlnvolvements Schulkultur – zentrales Element für eine gelingende Schulevaluation

Durchführungsobjektivität von EVALOG: Macht es einen Unterschied, ob in Kleinstbetrieben Arbeitgeber(innen) oder Kolleg(inn)en die Evaluierung psychischer Belastung durchführen?

Information & Service

Praxisberichte:

Wirkungsorientiertes Monitoring für die Gründungsförderprogramme > EXIST-Gründerstipendium vund > EXIST-Forschungstransfer v

Sich verändernde Rollenanforderungen an Akteurinnen und Akteure im Feld der Evaluation von Forschungs-, Technologie- und Innovationspolitik

Literaturüberblick und Rezensionen:

Wachter; Stockmann/Meyer; Hassnain/Kelly/Somma

DeGEval ...Info

WAXMANN

21. Jg. · Heft 1 · April 2022



Theorie, Methoden und Praxis der Evaluation

Zeitschrift für Evaluation 21. Jahrgang 2022, Heft 1, S. 55-75 https://doi.org/10.31244/zfe.2022.01.04 © 2022 Waxmann

Durchführungsobjektivität von EVALOG: Macht es einen Unterschied, ob in Kleinstbetrieben Arbeitgeber(innen) oder Kolleg(inn)en die Evaluierung psychischer Belastung durchführen?

Jochen Prümper,¹ Tom Schneeberg,² Alina Marie Prümper²

Zusammenfassung: In Kleinstbetrieben können Arbeitgeber(innen) die Arbeitsplatzevaluierung (Gefährdungsbeurteilung) psychischer Belastung auf zuverlässige und fachkundige Beschäftigte des eigenen Betriebes delegieren. In diesem Beitrag wird der Frage nachgegangen, inwieweit die wahrgenommene Gebrauchstauglichkeit, die Verständlich- und Handhabbarkeit bei der Durchführung einer Arbeitsplatzevaluierung, die emotionale Reaktion sowie die Beurteilung der psychischen Belastung selbst von dem Umstand beeinflusst wird, ob diese von Arbeitgeber(inne)n oder Kolleg(inn)en durchgeführt wird (Versuchsleiterartefakt).

An der Studie nahmen im Rahmen eines quasi-experimentellen Forschungsdesigns im Feld an 19 Arbeitsplatzevaluierungen in ebenso vielen Kleinstbetrieben 46 Proband(inn)en teil; davon 19 als Evaluierende (12 Arbeitgeber(innen) und sieben Kolleg(inn)en) sowie 27 als Beschäftigte. Die Ergebnisse zeigen: Beschäftigte beurteilen die Gebrauchstauglichkeit sowie die Verständlich- und Handhabbarkeit nicht signifikant unterschiedlich, aber sie zeigen positivere emotionale Reaktionen, wenn Kolleg(inn)en in der Rolle als Evaluierende das Verfahren durchführen.

Schlagwörter: Arbeitsplatzevaluierung, psychische Belastung, Gebrauchstauglichkeit, Versuchsleiterartefakt

Implementation Objectivity of EVALOG: Does it Make a Difference whether Employers or Colleagues Carry out the Evaluation of Mental Stress in Very Small Companies?

Abstract: In small companies, employers can delegate job evaluation (risk assessment) of psychological stress to reliable and competent employees of their own company. This article investigates the extent to which the perceived usability, understandability, and manageability in performing a job evaluation, the emotional response, and the assessment of the psychological stress itself, is influenced by the fact whether it is carried out by employers or colleagues (experimenter artifact/Rosenthal effect). As part of a quasi-experimental research design in the field, the study took part in 19 workplace evaluations of 19 small businesses. 46 subjects participated, 19 as evaluators (12 employers and seven colleagues) and 27 as employees. The results show: Employees do not assess the usability and dif-

¹ Hochschule f
ür Technik und Wirtschaft Berlin

² bao – Büro für Arbeits- und Organisationspsychologie GmbH

ficulty significantly differently, but they do show more positive emotional reactions when colleagues perform the procedure as evaluators.

Keywords: Job Evaluation, Psychological Stress, Usability, Rosenthal Effect

1. Hinführung zum Thema

Auch Arbeitgeber(innen) in Kleinstbetrieben (bis zu neun Beschäftigte) sind gesetzlich dazu verpflichtet, eine Gefährdungsbeurteilung (Arbeitsplatzevaluierung) psychischer Belastung durchzuführen. Die Durchführung selbst können sie auf zuverlässige und fachkundige Beschäftigte des eigenen Betriebes delegieren. Macht es für die Beurteilung der Gebrauchstauglichkeit, Verständlich- und Handhabbarkeit und emotionalen Reaktion, sowie der Beurteilung der psychischen Belastung selbst aus Beschäftigtensicht einen Unterschied, ob eine entsprechende Gefährdungsbeurteilung (Arbeitsplatzevaluierung) von dem bzw. der Arbeitgeber(in) selbst oder von einem Kollegen bzw. einer Kollegin durchgeführt wird?

2. Einleitung

Seit 2013 fordert das österreichische ArbeitnehmerInnenschutzgesetz (§ 4 ASchG) in der Arbeitsplatzevaluierung explizit die Berücksichtigung der psychischen Belastung. Kurz darauf erfolgte noch im gleichen Jahr eine entsprechende Novellierung des deutschen Arbeitsschutzgesetzes (§ 5 ArbSchG) für die Beurteilung der Arbeitsbedingungen.

Obwohl diese Novellierungen zu einem deutlichen Anstieg von Gefährdungsbeurteilungen (Arbeitsplatzevaluierungen) geführt haben, ist auffällig, dass selbige in Kleinstbetrieben deutlich seltener umgesetzt werden als in mittelgroßen oder großen Betrieben. Zudem ist der Anteil der Betriebe, die in diesem Zusammenhang auch psychische Belastung berücksichtigen, deutlich geringer als der Anteil der Betriebe, die die Durchführung von Gefährdungsbeurteilungen überhaupt angehen (vgl. Beck/Richter/Ertel/Morschhäuser 2012; GDA 2019). Gründe dafür werden u.a. in der Komplexität der rechtlichen Verpflichtungen, zusätzlicher Papierarbeit, monetären Belastungen sowie mangelnder Expertise und Unterstützung durch Spezialisten gesehen (vgl. EU-OSHA 2016). Zudem fällt es gerade Kleinstbetrieben schwer, eine Evaluierung psychischer Belastung sachgemäß durchzuführen, weil gängige Methoden wie Fragebögen, Gruppendiskussionen oder Beobachtungsinterviews vor allem für den Einsatz in größeren Organisationen entwickelt und erprobt wurden (vgl. Prümper/Strobach/Vowinkel 2018).

Forschungsergebnisse zeigen, dass die Durchführung einer Gefährdungsbeurteilung (Arbeitsplatzevaluierung) besonders erfolgreich ist, wenn Beschäftigte an diesem Prozess beteiligt werden (Nielsen/Randall/Holten/González 2010).

Vor diesem Hintergrund besteht ein besonderer Bedarf zur Etablierung von Verfahren zur Evaluierung (Gefährdungsbeurteilung) psychischer Belastung in Kleinstbetrieben, in denen den Beschäftigten selbst ein hohes Maß an Partizipation zuteil wird. Mit "EVALOG – Evaluierung psychischer Belastung im Dialog" entwickelten Prümper und Vowinkel (2019) im Auftrag der österreichischen Allgemeinen Unfallversicherungsanstalt (AUVA) ein derartiges Verfahren. Im Unterschied zu der von der AUVA entwickelten Arbeits-Bewertungs-Skala (ABS Gruppe) (vgl. Molnar/Prinkel/Friesenbichler 2013) kann EVALOG bereits ab einer bzw. einem Beschäftigten durchgeführt werden.

2.1 EVALOG – Evaluierung psychischer Belastung im Dialog

Bei der Evaluierung psychischer Belastung geht es darum, Belastung, die zu Fehlbeanspruchung führen kann, zu ermitteln und Maßnahmen zur gesundheitsförderlichen Gestaltung von Arbeitsbedingungen abzuleiten. Neben dem Einsatz in kleinen Betrieben eignet sich EVALOG auch zur Arbeitsplatzevaluierung sehr kleiner Tätigkeitsbereiche in größeren Organisationen.

Bislang wurde EVALOG u.a. bereits unter den Aspekten der Gebrauchstauglichkeit, Verständlich- und Handhabbarkeit, Teilnahmemotivation und emotionalen Reaktion aus Sicht von Evaluierenden und Beschäftigten (vgl. Prümper/Schneeberg/Prümper 2021) betrachtet oder es wurde der Frage nachgegangen, was Kleinstunternehmer(innen) überhaupt zur Teilnahme an der Entwicklung eines Verfahrens zur Evaluierung psychischer Belastung motiviert (vgl. Prümper/Strobach/Vowinkel 2018).

2.2 Qualitätsgrundsätze von EVALOG

Der Aufbau von EVALOG orientiert sich an den "Qualitätsgrundsätzen für Instrumente/Verfahren zur Gefährdungsbeurteilung psychischer Belastung" der GDA (2017). Hierzu wird in einem den Einsatz des Verfahrens begleitenden "Wegweiser" beschrieben, für welche Einsatzbereiche das Verfahren geeignet ist, wie Tätigkeiten und Ausführungsbedingungen berücksichtigt und wie Beschäftigte in den Prozess der Gefährdungsbeurteilung einbezogen werden können. Darüber hinaus beinhaltet der Wegweiser Methoden bzw. Hilfestellungen zur Beurteilung, ob Maßnahmen zur Minderung von Gefährdungen durch psychische Belastung erforderlich sind oder nicht. Das "Herzstück" von EVALOG ist der "KFZA - Kurz-Fragebogen zur Arbeitsanalyse" (Prümper/Hartmannsgruber/Frese 1995), der im Rahmen der Gefährdungsbeurteilung psychischer Belastung als "Standardinstrument" gilt (Appel/Schuler/Vogel/Oezelsel/Faller 2017: 1), "das keine zu hohen fachlichen Anforderungen an die Anwender stellt" (Hamborg/Schaper 2018: 51) und mit dem "reliabel und valide Problemfelder" (Uhle/Treier 2019: 382) erkannt werden können. Damit beruht EVALOG auf einem methodisch geprüften und ausgewiesenen Verfahren, welches alle relevanten Belastungsfaktoren berücksichtigt. Zur Sicherstellung, ob sich der

KFZA im Rahmen einer mehr oder weniger Einzelfalldiagnostik eignet, wurde der Einsatz bei Referenzbetrieben überprüft.

Der KFZA besteht aus den 11 Skalen und 26 Fragen und Aussagen (Items), die die psychische Belastung aus den Merkmalen der Arbeitstätigkeit erfassen. Die Items orientieren sich an der EN ISO 10075-1 (2018, Anhang A) in Verbindung mit ASchG § 7, Abs. 4a "§ 7 Grundsätze der Gefahrenverhütung" sowie ArbSchG § 4, Abs. 4 "§ 4 Allgemeine Grundsätze" und lassen sich den dort beschriebenen vier Komponenten psychischer Belastung aus der Arbeitssituation "Arbeitsaufgaben und Tätigkeiten" (z.B. "Vielseitigkeit"), "Sozial- und Organisationsklima" (z.B. "Soziale Unterstützung"), "Arbeitsumgebung" (z.B. "Umgebungsbelastung") sowie "Arbeitsabläufe und Arbeitsorganisation" (z.B. "Qualitative Arbeitsbelastung") zuordnen.

Die testtheoretische Qualität des KFZA wurde unter den Aspekten Reliabilität und Validität bereits in zahlreichen Studien nachgewiesen (vgl. u.a. Hartwig/Wirth/Bonin 2020; Hofmann/Stalder/Tschan/Häfeli 2014; Kada 2010; Keller/Meyer/Elfering/Semmer 2019; Li/Schaubroeck/Xie/Keller 2018; Mir et al. 2016; Rieder/Kraus/Vogl 2020). Ebenso wurde sein Einsatz in unterschiedlichen Erhebungsszenarien, wie z.B. einem interaktiven, elektronischen Votingsystem (EVS), unter den Aspekten der Objektivität, Zuverlässigkeit und Gebrauchstauglichkeit untersucht (vgl. Prümper/Schneeberg 2020).

2.3 Beschreibung von EVALOG

Bei EVALOG findet die Evaluierung im Dialog zwischen einem bzw. einer Evaluator(in) und ein bis drei Beschäftigten aus einer Tätigkeitsgruppe mit maximal fünf Arbeitnehmer(inne)n statt. Bei mehreren Tätigkeitsgruppen wird der Dialog entsprechend wiederholt. Neben dem Einsatz in Kleinstbetrieben kann das Verfahren auch in größeren Betrieben für Tätigkeitsgruppen bis zu max. fünf Beschäftigten zum Einsatz kommen.

Das Verfahren deckt alle sieben Evaluierungsphasen sowie die Dokumentation ab (vgl. Abbildung 1). Dabei bearbeiten die Evaluator(inn)en die schwarzen Teile 1 "Arbeitsbereiche und Tätigkeiten festlegen" sowie die Planung (z.B. durch das Festlegen verantwortlicher Personen) des 5. Schrittes "Maßnahmen umsetzen" im Allgemeinen allein und die graufarbenen Teile 2 "Psychische Belastung ermitteln", 3 "Psychische Belastung beurteilen", 4 "Maßnahmen festlegen", 6 "Wirksamkeit überprüfen" und 7 "Evaluierung anpassen" gemeinsam mit den Beschäftigten. Selbstverständlich können die Beschäftigten auch in allen sieben Schritten eingebunden werden und selbstredend werden die Evaluator(inn)en nicht immer diejenigen sein, die Maßnahmen "alleine" umsetzen; aber in ihrer Rolle als Arbeitgeber(in) (bzw. Beauftragte(r) des Betriebs) können sie auch Auftraggeber(in) sein.

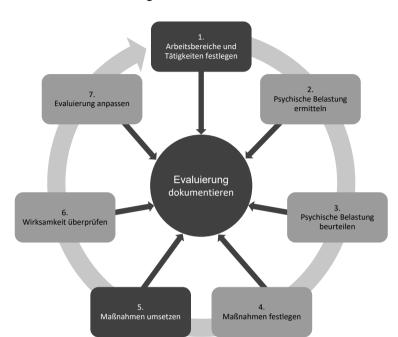


Abbildung 1: Sieben Schritte der Evaluierung (Gefährdungsbeurteilung) psychischer Belastung

EVALOG besteht aus fünf Teilen.

Der erste Teil "Evaluierung psychischer Belastung im Dialog" liefert einen Überblick über das Verfahren sowie eine detaillierte Erläuterung der sieben Umsetzungsschritte der Evaluierung (vgl. Abbildung 1). Zudem erfolgt eine kurze Einführung in die wichtigsten Themen rund um die Evaluierung und es wird – bereichert um Hinweise, Beispiele und Verweise auf Zusatzmaterial – eine praxisgerechte Anleitung zur Umsetzung der Evaluierung gegeben. Damit die Evaluator(inn)en gut auf den Evaluierungsdialog vorbereitet sind, ist die Lektüre dieses Teils des Wegweisers für sie unabdingbar.

Im zweiten Teil "Weiterführende Informationen zur Evaluierung" werden Hinweise zu Grundlagen einer guten Gesprächsführung (z.B. Mindeststandards, Grundhaltungen und Techniken), sowie dem arbeitspsychologischen Hintergrund (z.B. Belastungs-Beanspruchungs-Modell, persönlichkeitsförderliche Arbeitsgestaltung) und arbeitsrechtlichen Hintergrund (z.B. gesetzliche Anforderungen, Beteiligungsrechte) vermittelt.

Im dritten Teil "Kontaktstellen zur Unterstützung rund um die Evaluierung" finden sich Adressen zu Institutionen, die bei der Evaluierung und insbesondere bei psychischen Problemen im Betrieb bei Bedarf unterstützen.

Der vierte Teil "Informationsblätter" enthält Hinweise für Evaluator(inn)en und Arbeitnehmer(inn)en zum Hintergrund und zur Durchführung der Evaluierung. Diese Informationsblätter erleichtern den Betrieben die Umsetzung der Evaluierung und

können nach Bedarf kopiert oder ausgedruckt werden. Sie stehen zum Download bereit unter www.eval.at/evalog.

Der fünfte Teil enthält die "Arbeitsblätter", die zur Durchführung und Dokumentation der Evaluierung, entsprechend den gesetzlichen Anforderungen, benötigt werden und damit die Bearbeitung der "sieben Schritte der Evaluierung" (vgl. Abbildung 1) formal unterstützen.

2.4 Delegation im Arbeitsschutz

Gesundheit und Sicherheit ihrer Beschäftigten zu gewährleisten ist originäre Pflicht der Arbeitgeber(innen). Hierzu gehört auch die Durchführung der Gefährdungsbeurteilung (Arbeitsplatzevaluierung) psychischer Belastung. Allerdings haben Arbeitgeber(innen) auch die Möglichkeit, arbeitsschutzrechtliche Aufgaben auf Dritte, sogenannte "sonstige" Beauftragte (vgl. ArbSchG § 13, Abs 1, Ziff. 4), zu übertragen. In diesem Zusammenhang haben Arbeitgeber gemäß österreichischem ArbeitnehmerInnenschutzgesetz (§ 76 ASchG) u.a. die Pflicht, "in arbeitsphysiologischen, arbeitspsychologischen und sonstigen ergonomischen sowie arbeitshygienischen Fragen, insbesondere der Gestaltung der Arbeitsplätze und des Arbeitsablaufes" (§ 76 ASchG, Abs. 6), "bei der Ermittlung und Beurteilung der Gefahren" (§ 76 ASchG, Abs. 8) sowie "bei der Festlegung von Maßnahmen zur Gefahrenverhütung" (§ 76 ASchG, Abs. 9) "erforderlichenfalls weitere geeignete Fachleute hinzuzuziehen [...] insbesondere jedoch Arbeitspsychologen" (ASchG §4, Abs. 6).

Um einen qualitativ hochwertigen Prozess der psychischen Gefährdungsbeurteilung zu gewährleisten, müssen die Verantwortlichen über Kompetenzen in mehreren Bereichen verfügen. Gemäß "Positionspapier zu Gefährdungsbeurteilungen psychischer Belastungen: Anforderungen aus arbeits- und organisationspsychologischer Sicht" der Deutschen Gesellschaft für Psychologie (DGPs 2016) gehören hierzu 1. Kenntnisse über die Kriterien gut gestalteter Arbeit, 2. Kenntnisse der Analyseinstrumente und ihrer Einsatzmöglichkeiten, 3. Fähigkeit zur Ableitung von Arbeitsgestaltungsmaßnahmen, 4. Kompetenz zum Steuern komplexer Prozesse sowie 5. Evaluationskompetenz. In der "Zertifizierungsrichtlinie für Arbeits- und OrganisationspsychologInnen" des Berufsverbandes Österreichischer Psychologinnen und Psychologen (BÖP) ist die "Arbeitsplatzevaluierung des ASchG als Instrument zur Analyse, Bewertung, Gestaltung von betrieblichen Präventionsprozessen" verpflichtender Bestandteil (BÖP 2017: 11).

Auch Kleinstbetrieben wäre zu wünschen, dass sie die Evaluierung psychischer Belastung in die Hände einschlägig qualifizierter Arbeitspsycholog(inn)en legen könnten. Dieser Wunsch stößt in der betrieblichen Praxis aufgrund monetärer Beschränkungen jedoch häufig an seine Grenzen, so dass Arbeitgeber(innen) in Kleinstbetrieben im Allgemeinen genötigt sind, die gesetzlich verpflichtete Auflage "mit Bordmitteln" durch "zuverlässige und fachkundige Personen" zu organisieren, die die Aufgaben "in eigener Verantwortung" (ArbSchG § 13, Abs. 2) wahrnehmen. Von der notwendigen *Zuverlässigkeit* kann ausgegangen werden, wenn die Beauftragten aufgrund ihrer persönlichen Eigenschaften, ihres Verhaltens und ihrer Fä-

higkeiten zur ordnungsgemäßen Erfüllung der ihnen obliegenden Aufgaben geeignet sind; von der geforderten *Fachkunde*, wenn sie über das erforderliche theoretische Wissen, als auch über entsprechende praktische Fähigkeiten verfügen.

In der betrieblichen Praxis in Kleinstbetrieben kommen damit – wenn Arbeitgeber(innen) die Evaluierung nicht persönlich durchführen möchten – zur Delegation insbesondere Beschäftigte infrage, die dafür bekannt sind, ihre Aufgaben mit besonders gebotener Sorgfalt auszuführen, die bei ihren Kolleg(inn)en eine ausgesprochene Vertrauenswürdigkeit besitzen, über eine langjährige Berufserfahrung verfügen und sich nachdrücklich für das Thema Arbeitsschutz interessieren.

3. Forschungsfragen

Selbstverständlich sollte man sich die Frage stellen, ob es in einem arbeitsrechtlichen Kontext durch Abhängigkeiten zwischen Durchführenden und Teilnehmer(inne)n überhaupt eine Durchführungsobjektivität geben kann. Um dies jedoch weitestgehend sicherzustellen, ist EVALOG so aufgebaut, dass ein Teil der *Durchführungsobjektivität* durch die standardisierte Vorbereitung, Instruktionen und Gleichhaltung der Situationsbedingung (Gesprächssituation) sowie durch die zielgruppenorientierte, weiterführende Information gewährleistet wird. Die *Auswertungsobjektivität* wird durch verbale, äquidistant skalierte deskriptive Rating-Skalen und die *Interpretationsobjektivität* durch die Angabe von Grenzwerten, ab wann die Erarbeitung von Maßnahmen induziert ist, sichergestellt. Entsprechende standardisierte Instruktionen und Auswertungsregeln sind in dem Wegweiser beschrieben.

Da Arbeitgeber(innen) jedoch die Möglichkeit haben, den Evaluierungsdialog auf Beschäftigte zu übertragen, sollen mögliche "Versuchsleiterartefakte" (vgl. Eden 1990; Rosenthal/Fode 1963) als weitere mögliche kritische Einflussgröße im Rahmen der *Durchführungsobjektivität* (vgl. EN ISO 10075-3 2004) anhand von zwei Forschungsfragen überprüft werden.

Die *erste Forschungsfrage* lautet: Wird die von Arbeitnehmer(inne)n wahrgenommene Gebrauchstauglichkeit, emotionale Reaktion und Verständlichsowie Handhabbarkeit von EVALOG davon beeinflusst, ob Arbeitgeber(innen) oder Kolleg(inn)en die Evaluation durchführen?

Die zweite Forschungsfrage lautet: Kommen Arbeitgeber(innen) als Evaluator(inn)en zu anderen Einschätzungen der Gebrauchstauglichkeit, emotionalen Reaktion und wahrgenommenen Verständlich- und Handhabbarkeit (bezüglich der "weiterführenden Informationen zur Evaluierung" und der "Durchführung der einzelnen Schritte der Evaluierung") von EVALOG als Beschäftigte in der Rolle als Evaluator(in)?

4. Material und Methode

4.1 Forschungsdesign und Stichprobe

Die Gebrauchstauglichkeit des Wegweisers wurde hinsichtlich der ersten vier Schritte der Evaluierung (vgl. Abb. 1) mit österreichischen Kleinstbetrieben durch ein quasi-experimentelles Forschungsdesign im Feld im Zeitraum von April 2018 bis Juni 2019 geprüft. Die Akquisition der Kleinstbetriebe erfolgte mit Unterstützung der AUVA (Allgemeine Unfallversicherungsanstalt) sowie der WKO (Wirtschaftskammer Österreich) durch eine zufällige Auswahl³ aus allen Kleinstbetrieben mit einem bzw. einer bis neun Beschäftigten aus allen neun österreichischen Bundesländern aller sieben Wirtschaftskammern. Im Vorfeld wurden die Betriebe schriftlich über das Vorgehen und den Ablauf sowie über den Sinn, Zweck und Nutzen der Teilnahme informiert. Die Teilnahme war freiwillig und es erfolgte keine Aufwandsentschädigung. Nach Durchführung der Evaluierung beantworteten die Teilnehmenden zu den im Fokus stehenden Kriterien einen Fragebogen (wahlweise Papier/Bleistift oder Online).

Insgesamt umfasste die Stichprobe 19 Kleinstbetriebe mit durchschnittlich vier Beschäftigten (Min.=1, Max.=9) aus 14 (Min.=1, Max.=5) verschiedenen Branchen (z.B. Floristik, Tankstellen, Logistik) und sechs der neun Regionen Österreichs. In diesen 19 Betrieben wurden 19 Gefährdungsbeurteilungen (Arbeitsplatzevaluierungen) von 19 Evaluator(inn)en (davon 12 Arbeitgeber(innen) und sieben Beschäftigte) bei 27 Arbeitnehmer(inne)n (AN) durchgeführt (12 Evaluierungen mit 1 AN, 6 Evaluierungen mit 2 AN, 1 Evaluierung mit 3 AN). Die Kleinstbetriebe entschieden dabei eigenständig, ob die Evaluierung durch Arbeitgeber(innen) oder durch Kolleg(inn)en erfolgen sollte.

Von den 46 Proband(inn)en (19 Evaluator(inn)en und 27 Beschäftigten waren 28 Probanden weiblich und 15 männlich. Drei Teilnehmende erteilten keine Auskunft über ihr Geschlecht. Die Altersgruppe der 50- bis 59-Jährigen (n=13) machte den größten Teil der Proband(inn)en aus, gefolgt von der Gruppe der 40- bis 49-Jährigen (n=12), der Gruppe der 30- bis 39-Jährigen (n=7), der Gruppe der unter 30-Jährigen (n=6) sowie der Gruppe der über 60-Jährigen (n=4).

Nach Durchführung der Evaluierung beantworteten die Teilnehmenden (Evaluator(inn)en und Beschäftigte) einen Fragebogen zu den Kriterien. Dies erfolgte entweder unmittelbar im Anschluss an die Evaluierung als "Papier/Bleistift'-Befragung oder auf Wunsch online. Die Dauer des Ausfüllens der Befragung betrug ca. 45 Minuten.

³ Die Anzahl der kontaktierten Kleinstbetriebe wurde nicht dokumentiert.

4.2 Statistische Auswertung

Die Datenverarbeitung und -analyse wurde mittels des Softwarepakets SPSS Statistik 24 durchgeführt. Die Berechnung der Skalenmittelwerte erfolgte auf der Grundlage von mindestens zwei gültigen Angaben. Zur Überprüfung der *Durchführungs-objektivität* (also dem Vergleich der beiden Bedingungen 'Arbeitgeber(innen) als Evaluator(inn)en' vs. 'Kolleg(inn)en als Evaluator(inn)en') wurde der Mann-Whitney-U-Test für unabhängige Stichproben eingesetzt. Für alle Fragestellungen ergab die visuelle Überprüfung durch Q-Q-Plots, dass die Voraussetzungen für ein parametrisches Verfahren nicht erfüllt waren.

Des Weiteren wurde mittels Friedman-Test untersucht, ob es für die jeweilige Untersuchungsgruppe Unterschiede zwischen den zentralen Tendenzen der Unterkonstrukte der *Gebrauchstauglichkeit* (Effizienz, Effektivität, Zufriedenstellung) gab. Nachfolgend wurden Post-hoc-Tests durchgeführt, um zu bestimmen, welche Unterkonstrukte sich signifikant unterschieden. Um einen Fehler 1. Art beim multiplen Testen entgegenzuwirken, wurde die Bonferroni-Holm-Korrektur jeweils in Bezug auf eine Konstruktgruppe angewandt (vgl. Holm 1979). Zur Bestimmung der internen Konsistenzen wurden die Cronbachschen Alpha der vorliegenden Stichprobe berechnet. Fälle mit fehlenden Werten wurden ausgeschlossen.

4.3 Instrumente

4.3.1 Gebrauchstauglichkeit

Unter Gebrauchstauglichkeit wird das Ausmaß verstanden, "in dem ein Produkt durch bestimmte Benutzer in einem bestimmten Nutzungskontext genutzt werden kann, um bestimmte Ziele effektiv, effizient und zufriedenstellend zu erreichen" (EN ISO 10075-3 2004: 7). Gemäß EN ISO 10075-3 (2004: 12) müssen Messverfahren "effektiv und effizient sein und die Anforderungen der Anwender erfüllen, d.h. sie müssen bei minimalem Aufwand oder Mitteleinsatz das erwartete Präzisionsniveau im beabsichtigten Messbereich erreichen und zu den gewünschten Ergebnissen führen". Entsprechend wird in den internationalen Normen EN ISO 9241-11 (2018) und EN ISO 10075-3 (2004) das Konstrukt der Gebrauchstauglichkeit aufgegliedert in die Subkonstrukte Effektivität, Effizienz und Zufriedenstellung (vgl. EN ISO 9241-11 2018: 11) bzw. Zufriedenheit (vgl. EN ISO 10075-3 2004: 12). Da sich die Definitionen dieser drei Konstrukte wiederum recht komplexer Konzepte bedienen (s.u.), mussten zur Operationalisierung jeweils entsprechende Indikatoren gefunden werden. Hierzu generierten - im Sinne einer freien Assoziation - drei Probanden unabhängig voneinander für die jeweiligen Konzepte für sie passende Adjektive. Anschließend einigten sie sich - zur Messbarmachung der drei Subkonstrukte – auf die vier ihres Erachtens treffendsten Adjektive. Zur differenzierten Beantwortung der so generierten Items stand jeweils eine 5-stufige Ratingskala von

1=,stimme nicht zu' bis 5=,stimme sehr zu' zur Verfügung. Die interne Konsistenz der gesamten Skala *Gebrauchstauglichkeit* betrug α =,86.

4.3.1.1 Effektivität

EN ISO 9241-11 (2018: 11) versteht unter *Effektivität* "Genauigkeit und Vollständigkeit, mit denen Benutzer bestimmte Ziele erreichen". Die Subskala *Effektivität* wurde durch folgende Items abgedeckt: "EVALOG ist eine ... (1) effektive, (2) überzeugende, (3) schlüssige und (4) plausible Methode zur Evaluierung psychischer Belastungen in der Arbeitssituation". Die interne Konsistenz der Subskala *Effektivität* betrug α =,84. Die signifikanten (p<0.01) Korrelationen nach Pearson mit der Subskala *Effizienz* betrug r=,74 und mit der Subskala *Zufriedenstellung* r=.56.

4.3.1.2 Effizienz

EN ISO 9241-11 (2018: 11) definiert die *Effizienz* eines Messverfahrens als ein relatives Konzept, abhängig von seiner Effektivität (z.B. dem Präzisionsniveau) und dem Aufwand, der zur Durchführung, Analyse und Interpretation der Messung und ihrer Ergebnisse erforderlich ist". Die Subskala *Effizienz* wurde durch folgende Items abgedeckt: "EVALOG ist eine ... (1) effiziente, (2) zeitsparende, (3) praktikable und (4) griffige Methode zur Evaluierung psychischer Belastungen in der Arbeitssituation". Die interne Konsistenz der Subskala *Effizienz* betrug α =,81. Die signifikanten (p<0,01) Korrelationen nach Pearson mit der Subskala *Effektivität* betrug r=,74 und mit der Subskala *Zufriedenstellung* r=,48.

4.3.1.3 Zufriedenstellung

EN ISO 9241-11 (2018: 11) beschreibt *Zufriedenstellung* als das "Ausmaß der Übereinstimmung der physischen, kognitiven und emotionalen Reaktionen des Benutzers, die aus der Benutzung eines Systems, eines Produkts oder einer Dienstleistung resultieren, mit den Benutzererfordernissen und Benutzererwartungen". Die Subskala *Zufriedenstellung* bestand aus den Items: "EVALOG ... (1) verdeutlicht in der Einleitung zufriedenstellend ihren Nutzen, (2) macht einen soliden wissenschaftlichen Eindruck, (3) vermittelt glaubwürdig, dass die erhobenen Daten vertraulich behandelt werden und (4) lässt sich problemlos bearbeiten". Die interne Konsistenz der Subskala *Zufriedenstellung* betrug α =,51. Die signifikanten (p<0,01) Korrelationen nach Pearson mit der Skala *Effektivität* betrug r=,56 und mit der Skala *Effizienz* r=,48.

4.3.2 Emotionale Reaktion

Die *emotionale Reaktion* schätzten die Teilnehmer(innen) mit einer Ratingskala von 1 (negativer Pol) bis 5 (positiver Pol) auf einem konzipierten semantischen Differential ein. Die Skala bestand aus den Items: "Wie fühlen Sie sich nach der Evalu-

ierung im Dialog? (1) entspannt – angespannt, (2) zufrieden – unzufrieden und (3) voll Energie – schlapp". Die interne Konsistenz der Skala *Emotionale Reaktion* betrug α =,85.

4.3.4 Verständlich- und Handhabbarkeit

Bei der Beurteilung der *Verständlichkeit* geht es um die Eigenschaft von Texten – im konkreten Fall die der "Weiterführenden Informationen zur Evaluierung" hinsichtlich (1) der "Gesprächsführung", (2) des "Arbeitspsychologischen Hintergrundes" sowie (3) des "Arbeitsrechtlichen Hintergrundes" – bzgl. Leserlichkeit, Lesbarkeit und weiteren Merkmalen wie Gegliedertheit und Stimulanz (interne Konsistenz der Skala *Verständlichkeit* α =,62).

Bei der Beurteilung der *Handhabbarkeit* handelt es sich um die Eigenschaft eines Gegenstandes, Werkzeuges, Apparates – im konkreten Fall der "Arbeitsblätter" die zur Durchführung und Dokumentation der Evaluierung zum (4) "Festlegen von Arbeitsbereichen", (5) "Ermitteln psychischer Belastung", (6) "Beurteilen psychischer Belastung" und (7) "Festlegen von Maßnahmen" benötigt werden – (mit der Hand) gebraucht werden zu können (interne Konsistenz der Skala *Handhabbarkeit* α =,46); die Schritte 5 bis 7 der Evaluierung konnten zum aktuellen Zeitpunkt noch nicht einer Überprüfung unterzogen werden.

5. Ergebnisse

5.1 Erste Forschungsfrage – Sicht der Beschäftigten

Wird die von Arbeitnehmer(innen) wahrgenommene Gebrauchstauglichkeit, emotionale Reaktion und wahrgenommene Verständlich- und Handhabbarkeit bei der Durchführung von EVALOG davon beeinflusst, ob Arbeitgeber(innen) oder Kolleg(inn)en die Evaluation durchführen?

5.1.1 Gebrauchstauglichkeit

Die Gebrauchstauglichkeit gesamt bewerteten beide Gruppen zusammen mit MW=4,1 (SD=,62) und die Unterkonstrukte Effektivität MW=4,01 (SD=,68), Effizienz MW=4,06 (SD=,65) und Zufriedenstellung MW=4,34 (SD=,44). Die Gebrauchstauglichkeit von EVALOG wurde von den Beschäftigten der beiden Bedingungen ('Arbeitgeber(in) als Evaluator(in)' vs. 'Kollege bzw. Kollegin als Evaluator(in)') ähnlich gut bewertet (siehe Tabelle 1). Es bestanden keine signifikanten Unterschiede zwischen den beiden Gruppen.

		Sich	nt der Be	eschäfti	Test					
	Arbeitgeber(in) als Evaluator(in)				ge bzw. K Evaluato	_	Mann-Whitney-U			
	N	MW	SD	N	MW	SD	z	U	p	p ^{kor}
Gebrauchstaug- lichkeit (gesamt)	18	4,20	,63	7	3,86	,57	-1,25	42,50	,22	,57
Effektivität	18	4,18	,66	7	3,82	,72	-1,35	41,00	,19	,57
Effizienz	18	4,18	,67	7	3,75	,52	-1,59	37,00	,11	,44
Zufriedenstel- lung	18	4,33	,48	7	4,36	,35	-,19	60,00	,87	,87

Tabelle 1: Gebrauchstauglichkeit (Vergleich der beiden Bedingungen aus Sicht der Beschäftigten)

Anmerkungen: MW=Mittelwert; SD=Standardabweichung; z=Z-Wert; U=U-Wert; p=exakte Signifikanz; p^{kor} =p-Werte nach Bonferroni-Holm-Korrektur; Skalierung: 1=,stimme nicht zu' bis 5=,stimme sehr zu'

Die Unterkonstrukte der Gebrauchstauglichkeit unterschieden sich bei der Evaluierung durch Arbeitgeber(innen) (Friedman-Test: $\chi^2(2)$ =,25; p=,89; n=18) nicht signifikant untereinander. Bei der Evaluierung durch Kolleg(inn)en unterschieden sich die Unterkonstrukte (Friedman-Test: $\chi^2(2)$ =7,52; p=,02; n=7) signifikant untereinander. Nachfolgende Post-hoc-Tests (Dunn-Bonferroni-Tests) ergaben, dass die *Zufriedenstellung* bei der Evaluierung durch Kolleg(inn)en von den Mitarbeiter(inne)n signifikant besser als die *Effizienz* (z=-1,21; $p^{angepasst}$ =,07; Effektstärke nach Cohen (1992) mit starkem Effekt: d=1.38) mit einer marginalen Signifikanz von p=,10 bewertet wurde.

5.1.2 Emotionale Reaktion

Die emotionale Reaktion hat sich bei beiden Gruppen zusammen positiv dargestellt MW=3,97 (SD=,97). Beschäftigte, deren psychische Belastung von einer Kollegin oder einem Kollegen ("Kollege bzw. Kollegin als Evaluator(in)") evaluiert wurde, reagierten emotional signifikant positiver auf den Evaluierungsdialog (MW=4,52; SD=,47; exakter Mann-Whitney-U-Test: U=31,00; p=,05) als Beschäftigte (MW=3,61; SD=1.00), deren psychische Belastung von einer bzw. einem Arbeitgeber(in) betrachtet wurde (siehe Tabelle 2). Die Effektstärke nach Cohen (1992) lag bei d=1,12, was einem starken Effekt entspricht.

		Si	cht der B	eschäftig	ten		Test			
	Arbeitgeber(in) als Evaluator(in)				e bzw. Koll Evaluator(ii		Mann-Whitney-U			
	N	MW	SD	N	MW	SD	z	U	p	
Emotionale Reaktion	18	3,61	1,00	7	4,52	,47	-1,96	31,00	,05	

Tabelle 2: Emotionale Reaktion (Vergleich der beiden Bedingungen aus Sicht der der Beschäftigten)

Anmerkungen: MW=Mittelwert; SD=Standardabweichung; z=Z-Wert; U=U-Wert; p=exakte Signifikanz; Skalierung: 1=,negative emotionale Reaktion' bis 5=,positive emotionale Reaktion'

5.1.3 Verständlich- und Handhabbarkeit

Die Verständlichkeit wurde von beiden Gruppen zusammen mit MW=2,07 (SD=1,02) und die Handhabbarkeit mit MW=2,18 (SD=1,12) insgesamt als leicht eingestuft. Die beiden Beschäftigtengruppen unterschieden sich weder signifikant hinsichtlich der Verständlichkeit (siehe Tabelle 3) noch hinsichtlich der Handhabbarkeit (siehe Tabelle 4). Bezüglich der Verständlich- und Handhabbarkeit macht es also keinen Unterschied, ob EVALOG durch die bzw. den Arbeitgeber(in) oder eine(n) Kollegin oder Kollegen durchgeführt wurde.

Tabelle 3: Verständlichkeit (Vergleich der Bedingungen aus Sicht der Beschäftigten)

		Sicl	nt der B	eschäfti		Test				
	Arbeitgeber(in) als Evaluator(in)			Kollege bzw. Kollegin als Evaluator(in)			Mann-Whitney-U			
	N	MW	SD	N	MW	SD	Z	U	p	p ^{kor}
Verständlichkeit (gesamt)	13	2,38	1,10	7	1,48	,47	-2,13	19,00	,04	,16
Gesprächs- führung	13	2,23	1,24	7	1,14	,38	-1,17	31,50	,27	,44
Arbeitspsycho- logischer Hintergrund	18	1,94	1,43	7	1,14	,38	-1,32	45,00	,22	,44
Arbeitsrechtlicher Hintergrund	13	2,77	1,36	6	1,83	,75	-1,50	22,50	,14	,42

Anmerkungen: MW=Mittelwert; SD=Standardabweichung; z=Z-Wert; U=U-Wert; p=exakte Signifikanz; p^{kor} =p-Werte nach Bonferroni-Holm-Korrektur; Skalierung: 1=,sehr leicht' bis 5=,sehr schwierig/kompliziert'

Die Items der *Verständlichkeit* zur Evaluierung (siehe Tabelle 4) unterschieden sich untereinander nicht signifikant in der Bedingung 'Arbeitgeber(in) als Evaluator(in)' (Friedman-Test: $\chi^2(2)=2,39$. Bei der Bedingung 'Kollege bzw. Kollegin als Evaluator(in)' (Friedman-Test: $\chi^2(2)=5,60$; p=,06; n=6) unterschieden sich die Items ebenfalls – aber knapp – nicht signifikant.

Tabelle 4: Handhabbarkeit (Vergleich der Bedingungen aus Sicht der Beschäftigten)

		Sic	ht der Be	eschäfti		Test				
	Arbeitgeber(in) als Evaluator(in)			Kollege bzw. Kollegin als Evaluator(in)			Mann-Whitney-U			
	N	MW	SD	N	MW	SD	z	U	р	p ^{kor}
Handhabbarkeit (gesamt)	18	2.23	1,26	7	2,04	,67	03	62,50	,99	,99
Festlegen von Arbeitsbereichen	/	1	1	/	1	/	1	1	1	1
Ermitteln psychischer Belastung	18	2,28	1,49	7	1,86	,69	-0,25	59,00	,79	,99
Beurteilen psychi- scher Belastung	18	2,22	1,40	7	1,71	,95	-0,75	51,50	,51	,99
Festlegen von Maß- nahmen	17	2,29	1,16	6	3,00	1,26	-1,06	36,50	,33	,99

Anmerkungen: MW=Mittelwert; SD=Standardabweichung; z=Z-Wert; U=U-Wert; p=exakte Signifi-kanz; p^{kor}=p-Werte nach Bonferroni-Holm-Korrektur; Skalierung: 1 =, sehr leicht' bis 5=, sehr schwierig/kompliziert'. Das "Festlegen von Arbeitsbereichen" oblag lediglich den Evaluator(inn)en und wurde somit nicht von den Beschäftigten beurteilt.

Die Items zur *Handhabbarkeit* (siehe Tabelle 4) unterschieden sich untereinander nicht signifikant in der Bedingung "Arbeitgeber(in) als Evaluator(in)" (Friedman-Test: $\chi^2(2)=,08$; p=,96; n=16). In der Bedingung "Kollege bzw. Kollegin als Evaluator(in)" unterschieden sich die Items (Friedman-Test: $\chi^2(2)=4,77$; p=,09; n=6) auf einem Signifikanzniveau von p=,05 ebenfalls nicht.

5.2 Zweite Forschungsfrage – Sicht der Evaluierer(innen)

Kommen Arbeitgeber(innen) als Evaluator(inn)en zu anderen Einschätzungen der Gebrauchstauglichkeit, emotionalen Reaktion und wahrgenommenen Verständlichund Handhabbarkeit bei der Durchführung von EVALOG als Beschäftigte in der Rolle als Evaluator(in)?

5.2.1 Gebrauchstauglichkeit

Die Ergebnisse zur *Gebrauchstauglichkeit* (siehe Tabelle 5) zeigen, dass beide Gruppen die Gebrauchstauglichkeit sehr positiv bewerteten. Selbige betrug – für beide Gruppen zusammen – für die *Gebrauchstauglichkeit gesamt MW*=4,52 (*SD*=,47) und für die Unterkonstrukte *Effektivität MW*=4,47 (*SD*=,60), *Effizienz MW*=4,53 (*SD*=,49) und *Zufriedenstellung MW*=4,61 (*SD*=,44). Arbeitgeber(innen) und Beschäftigte in der Rolle des Evaluators bzw. der Evaluatorin unterschieden sich weder in ihrer Beurteilung der Gebrauchstauglichkeit gesamt noch hinsichtlich der Beurteilung der zugehörigen drei Unterkonstrukte.

Tabelle 5: Gebrauchstauglichkeit (Vergleich der Bedingungen von Evaluator(inn)en)

		Sicht	Test								
	Arbeitgeber(innen) als Evaluator(inn)en			l .	Beschäftigte als Evaluator(inn)en			Mann-Whitney-U			
	N	MW	SD	N	MW	SD	Z	U	р	p ^{kor}	
Gebrauchstaug- lichkeit (gesamt)	12	4,60	,50	7	4,38	,50	-,86	32,00	,42	,99	
Effektivität	12	4,58	,64	7	4,29	,51	-1,34	27,00	,19	,76	
Effizienz	12	4,58	,51	7	4,43	,47	-,79	33,00	,49	,99	
Zufriedenstellung	12	4,67	,43	7	4,50	,48	-,79	33,00	,45	,99	

Anmerkungen: MW=Mittelwert; SD=Standardabweichung; z=Z-Wert; U=U-Wert; p=exakte Signifikanz; p^{kor} =p-Werte nach Bonferroni-Holm-Korrektur; Skalierung: 1=,stimme nicht zu' bis 5=,stimme sehr zu'

Die Unterkonstrukte der Gebrauchstauglichkeit untereinander unterschieden sich weder bei den Arbeitgeber(inne)n (Friedman-Test: $\chi^2(2)=1.81$; p=,41; n=12) noch bei den Beschäftigten als Evaluator(inn)en (Friedman-Test: $\chi^2(2)=2,71$; p=,26; n=7) signifikant.

5.2.2 Emotionale Reaktion

Die *emotionale Reaktion* betrug für beide Gruppen zusammen *MW*=4,44 (*SD*=,58) und stellte sich von beiden Gruppen (siehe Tabelle 6) auf einem ähnlich positiven Niveau dar. Es zeigte sich kein signifikanter Unterscheid zwischen den Gruppen.

		Sich	Test							
	Arbeitgeber(innen) als Evaluator(inn)en				schäftigte aluator(inn		Mann-Whitney-U			
	N	MW	SD	N	MW	SD	z	U	р	
Emotionale Reaktion	11	4,61	,42	7	4,19	,74	-1,16	26,00	,25	

Tabelle 6: Emotionale Reaktion (Vergleich der Bedingungen von Evaluator(inn)en)

Anmerkungen: MW=Mittelwert; SD=Standardabweichung; z=Z-Wert; U=U-Wert; p=exakte Signifikanz; Skalierung: 1=,negative emotionale Reaktion' bis 5=,positive emotionale Reaktion'

5.2.4 Verständlich- und Handhabbarkeit

Die *Verständlichkeit* wurde von beiden Gruppen zusammen mit *MW*=2,00 (*SD*=,46) und die *Handhabbarkeit* mit *MW*=1,77 (*SD*=,57) als leicht bewertet. Die Arbeitgeber(innen) als Evaluator(inn)en beurteilten das Verfahren hinsichtlich der *Verständlichkeit* (siehe Tabelle 7) als auch bezüglich der *Handhabbarkeit* (Tabelle 8) auf einem gleich guten Niveau wie die Beschäftigten als Evaluator(inn)en.

Tabelle 7:	Verständlichkeit	(Vergleich der Bedingungen von Evaluator(inn)	en)

		Sicht	der Eva	aluator(i		Test				
	Arbeitgeber(innen) als Evaluator(inn)en			Beschäftigte als Evaluator(inn)en			Mann-Whitney-U			
	N	MW	SD	N	MW	SD	z	U	р	p ^{kor}
Verständlichkeit (gesamt)	10	1,90	,39	7	2,14	,54	-1,01	25,00	,38	,69
Gesprächs- führung	10	1,80	,63	7	2,29	,49	-1,61	21,00	,23	,69
Arbeitspsycho- logischer Hintergrund	11	1,45	,69	7	2,14	,69	-1,96	18,50	,09	,36
Arbeitsrechtlicher Hintergrund	10	2,40	,84	7	2,00	,82	-1,05	25,00	,33	,69

Anmerkungen: MW=Mittelwert; SD=Standardabweichung; z=Z-Wert; U=U-Wert; p=exakte Signifikanz; p^{kor} =p-Werte nach Bonferroni-Holm-Korrektur; Skalierung: 1=,sehr leicht' bis 5=,sehr schwierig/kompliziert'

Die Items der *Verständlichkeit* (siehe Tabelle 8) unterschieden sich untereinander weder in der Bedingung 'Arbeitgeber(innen) als Evaluator(inn)en' (Friedman-Test: $\chi^2(2)=3,93$; p=,14; n=10) noch in der Bedingung 'Beschäftigte als Evaluator(inn)en' (Friedman-Test: $\chi^2(2)=1,20$; p=,55; n=7).

		Sicht	der Eva	Test						
	Arbeitgeber(innen) als Evaluator(inn)en			Beschäftigte als Evaluator(inn)en			Mann-Whitney-U			
	N	MW	SD	N	MW	SD	z	U	р	p ^{kor}
Handhabbarkeit (gesamt)	12	1,64	,52	7	2,00	.61	-1.45	25,00	,17	,68
Festlegen von Arbeitsbereichen	11	1,27	,47	6	1,83	1,17	-1,08	24,00	,43	,99
Ermitteln psychi- scher Belastung	12	1,92	,79	7	2,00	,58	-0,28	39,00	,85	,99
Beurteilen psychi- scher Belastung	12	1,83	,83	7	2,00	,58	-0,55	36,00	,59	,99
Festlegen von Maßnahmen	11	1,55	1,04	7	2,29	,95	-1,61	22,50	,12	,60

Tabelle 8: Handhabbarkeit (Vergleich der Bedingungen von Evaluator(inn)en)

Anmerkungen: MW=Mittelwert; SD=Standardabweichung; z=Z-Wert; U=U-Wert; p=exakte Signifikanz; p*or=p-Werte nach Bonferroni-Holm-Korrektur; Skalierung: 1=,sehr leicht* bis 5=,sehr schwierig/kompliziert*

Die Items zur *Handhabbarkeit* (siehe Tabelle 8) unterschieden sich untereinander weder in der Bedingung 'Arbeitgeber(in) als Evaluator(in)' (Friedman-Test: $\chi^2(2)=3,63$; p=,31; n=10) noch in der Bedingung 'Evaluierer(in) = Kolleg(in)' (Friedman-Test: $\chi^2(2)=3,40$; p=,33; n=6).

6. Diskussion

Die Ergebnisse der Untersuchung zeigen, dass die Beschäftigten – unabhängig davon, ob Arbeitgeber(innen) oder Kolleg(inn)en als Evaluator(inn)en fungieren – die Gebrauchstauglichkeit von EVALOG als gut bewerten. Die Effektivität und Effizienz werden dabei tendenziell von Beschäftigten der Bedingung 'Arbeitgeber(in) als Evaluator(in)' – besser als in der Bedingung 'Kollege bzw. Kollegin als Evaluator(in)' bewertet. Dieser visuelle Effekt kann auf Stichprobengrößenunterschiede sowie auf Effekte der Interviewsituation (z.B. soziale Erwünschtheit) zurückzuführen sein. Zudem reagieren die Beschäftigten recht positiv auf den Evaluierungsdialog und sie stufen die Verständlich- und Handhabbarkeit als gut ein. Hervorzuheben ist in diesem Zusammenhang, dass Beschäftigte, deren psychische Belastungen von einer Kollegin bzw. einem Kollegen evaluiert werden, emotional positiver (entspannter, zufriedener und energievoller) auf den Evaluierungsdialog reagieren als die Beschäftigten, die den Evaluierungsdialog mit ihrem bzw. ihrer Arbeitgeber(in) durchführen. Aus Beschäftigtensicht handelt es sich damit bei dem Evaluierungsdialog mittels EVALOG um ein gebrauchstaugliches und einfach anzu-

wendendes Verfahren, welches gerne von den Arbeitgeber(inne)n zum Einsatz gebracht werden kann, bei dem sich die Beschäftigten jedoch noch etwas mehr wohl fühlen, wenn selbiges in die Hände ihrer Kolleg(inn)en gelegt wird.

Allerdings muss angemerkt werden, dass aufgrund der Durchführungsbedingungen von EVALOG und dem Studiendesign (z.B. keine Doppel-Verblindung) – wenngleich im Leitfaden der Hinweis zur Grundhaltung der Gesprächsführung (z.B. Repressionsfreiheit) vermittelt wird – Probandeneffekte sowie sozial erwünschtes Antwortverhalten auch durch Beeinflussung der Evaluator(inn)en nicht komplett ausgeschlossen werden können. Das Review von Wegge, Shemla und Haslam (2014) zeigt auf, dass Führungskräfte einen großen Einfluss auf das Gruppenklima und die gemeinsame soziale Identität haben, was sich beispielsweise in der kollektiven Bewertung und im Umgang mit psychischer Belastung oder im Vertrauensverhältnis zeigen kann. Bisher wurden die internen Evaluator(inn)en ausschließlich durch die Hinweise im Leitfaden auf ihre moderativen Fähigkeiten als Qualitätsmerkmal hin unterstützt. Ein Vorbereitungstraining für Evaluator(inn)en – in dem der Umgang mit Tabus, der Aufbau von Vertrauen oder Moderations- und Kommunikationstechniken reflektiert werden – wäre wünschenswert, ist aber in der Praxis von den Kleinstbetrieben alleine nur schwer realisierbar.

Des Weiteren wurde nicht überprüft, ob Expert(inn)en zu anderen Ergebnissen als interne Evaluator(inn)en kämen; diesbezüglich bedarf es weiterer Studien. Unter der Annahme, dass Unterschiede bestehen, könnte in Zukunft untersucht werden, ob ein Vorbereitungstraining für interne Evaluator(inn)en zu einer Verbesserung hinsichtlich der Angleichung an die Ergebnisse von Expert(inn)en ('Goldstandard') führen würde.

Die Datengrundlage bezüglich der Anzahl an österreichischen Kleinstbetrieben mit einer durchschnittlichen Dienstnehmeranzahl von 1 bis 9 (*N*=9 von ca. 280.000; Stand: 2020) und der dortigen Anzahl von Beschäftigten (*N*=46 von ca. 724.000; Stand: 2020) sowie der Anzahl der Fachverbände bzw. Innungsgruppen (*N*=13 von 93; Stand: 2020) fällt zu gering aus, um die Ergebnisse generalisieren zu können. Weitere Einschränkungen der Datenrepräsentativität betreffen Effekte der Selbstselektion. Es kann nicht ausgeschlossen werden, ob die freiwillige Teilnahme der Kleinstbetriebe auf entsprechenden betrieblichen Bedürfnissen beruhte. Diesbezüglich wären weitere (Doppelblind-)Studien an größeren Stichproben wünschenswert.

Bezüglich der Operationalisierung zu den Subskalen der Gebrauchstauglichkeit muss angemerkt werden, dass eine hohe Korrelation zwischen der Subskala der Effektivität und der Effizienz von r=,74 bei dieser Stichprobe besteht. Des Weiteren weisen die Skalen Zufriedenstellung und Handhabbarkeit eine recht geringe interne Konsistenz auf.

Neben den Beschäftigten beurteilen auch beide Evaluator(inn)engruppen die Gebrauchstauglichkeit von EVALOG als äußerst positiv und die Verständlich- und Handhabbarkeit als gut. Ebenso stellt sich die emotionale Reaktion auf den Evaluierungsdialog für beide Gruppen auf einem ähnlich positiven Niveau dar. Auch aus Sicht von Evaluator(inn)en handelt es sich bei EVALOG um ein gebrauchstaugliches und einfach anzuwendendes Verfahren, bei dessen Einsatz sie sich wohlfühlen.

Eine weitere Forschungsfrage, die einem möglichen Effekt nachgeht, ob es hinsichtlich der Beurteilung der psychischen Belastung einen Unterschied macht, ob diese von Arbeitgeber(inne)n oder von Kolleg(inn)en als Evaluator(inn)en durchgeführt wird, ist denkbar, aber mit dem in dieser Studie durchgeführten Forschungsdesign nicht umsetzbar. Hierfür wären dann die Evaluierenden sowie die beurteilten Arbeitsplätze und -situationen vor Ort als Varianzquelle zu beachten.

Zusammenfassend zeigen die Ergebnisse, dass im Großen und Ganzen kein Versuchsleiterartefakt für den Evaluierungsdialog befürchtet werden muss. Allerdings sollte durch weitere Studien untersucht werden, wie sich die Durchführungsobjektivität von EVALOG gemäß EN ISO 10075-3 (2004) noch weiter optimieren lässt. Wünschenswert wäre in diesem Zusammenhang ein gezieltes Vorbereitungsprogramm (und dessen Evaluierung), welches über die WKO in Kooperation mit der AUVA den Kleinstbetrieben der jeweiligen Innungsgruppen angeboten wird, so dass gleichzeitig die Reflexion branchenspezifischer psychischer Belastungen Bestandteil der Qualifizierung sind.

In der Praxis sprechen die Ergebnisse zudem dafür, dass Arbeitgeber(innen) beruhigt von ihrem Delegationsrecht Gebrauch machen und den Evaluierungsdialog psychischer Belastung an einen "Beschäftigten des Vertrauens" übertragen können.

Literatur

- Appel, Patricia/Schuler, Michael/Vogel, Heiner/Oezelsel, Amina/Faller, Hermann (2017): Short Questionnaire for Workplace Analysis (KFZA): Factorial Validation in Physicians and Nurses Working in Hospital Settings. In: Journal of Occupational Medicine and Toxicology, 12 (11), S. 1-10. https://doi.org/10.1186/s12995-017-0157-6
- Beck, David/Richter, Gabriele/Ertel, Michael/Morschhäuser, Martina (2012): Gefährdungsbeurteilung bei psychischen Belastungen in Deutschland. In: Prävention und Gesundheitsförderung, 7 (2), S. 115-119. https://doi.org/10.1007/s11553-011-0326-x
- BÖP Berufsverband Österreichischer PsychologInnen (2017): Zertifizierungsrichtlinie für Arbeitsund OrganisationspsychologInnen. Verfügbar unter: https://www.aap.co.at/akademie/AO-Psychologinnen Zertifizierungsrichtlinie v5 Juli 2016.pdf [01.02.2021].
- Cohen, Jacob (1992): A Power Primer. In: Psychological Bulletin, 112 (1), S. 155-159. https://doi.org/10.1037/0033-2909.112.1.155
- DGPs Deutsche Gesellschaft für Psychologie (2016): Positionspapier zu Gefährdungsbeurteilungen psychischer Belastungen: Anforderungen aus arbeits- und organisationspsychologischer Sicht. Verfügbar unter: https://dgps.de/index.php?id=2000498&tx_ttnews%5Btt_news%5D=1738&cHash=d02dfd6aff604ed6401957c55f0723da. [01.02.2021].
- Eden, Dov (1990): Pygmalion in Management: Productivity as A Self-Fulfilling prophecy. New York: The Free Press.
- EN ISO 9241-11 (2018): Ergonomie der Mensch-System-Interaktion Teil 11: Gebrauchstauglichkeit: Begriffe und Konzepte. Berlin: Beuth.
- EN ISO 10075-1 (2018): Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung Teil 1: Allgemeine Aspekte und Konzepte und Begriffe. Berlin: Beuth.
- EN ISO 10075-3 (2004): Ergonomische Grundlagen bezüglich psychischer Arbeitsbelastung Teil 3: Grundsätze und Anforderungen an Verfahren zur Messung und Erfassung psychischer Arbeitsbelastung. Berlin: Beuth.
- EU-OSHA European Agency for Safety and Health at Work (2016): Second European Survey of Enterprises on New and Emerging Risks (ESENER-2) Overview Report: Managing Safety and Health at Work. Luxembourg: Publications Office of the European Union.

- GDA (Gemeinsame Deutsche Arbeitsschutzstrategie) (2017): Arbeitsschutz in der Praxis. Empfehlungen zur Umsetzung der Gefährdungsbeurteilung psychischer Belastung. Berlin: GDA.
- GDA (Gemeinsame Deutsche Arbeitsschutzstrategie) (2019): Arbeitsschutz auf dem Prüfstand. Abschlussbericht zur Dachevaluation der Gemeinsamen Deutschen Arbeitsschutzstrategie. Berlin: GDA.
- Hamborg, Kai-Christoph/Schaper, Niclas (2018): Analyse, Bewertung und Gestaltung von Aufgaben, Arbeitstätigkeiten und Arbeitssystemen. In: Greif, Siegfried/Hamborg, Kai-Christoph (Hg.): Methoden der Arbeits-, Organisations- und Wirtschaftspsychologie, Enzyklopädie der Psychologie, Band B/III/3. Göttingen: Hogrefe, S. 1-79.
- Hartwig, Matthias/Wirth, Manuel/Bonin, Dominik (2020): Insights about Mental Health Aspects at Intralogistics Workplaces A Field Study. In: International Journal of Industrial Ergonomics, 76 (4), S. 1-7. https://doi.org/10.1016/j.ergon.2020.102944
- Hofmann, Claudia/Stalder, Barbara/Tschan, Franziska/Häfeli, Kurt (2014): Support from Teachers and Trainers in Vocational Education and Training: The Pathways to Career Aspirations and Further Career Development. In: International Journal for Research in Vocational Education and Training, 1 (1), S. 1-20. https://doi.org/10.13152/IJRVET.1.1.5
- Holm, Sture (1979): A Simple Sequentially Rejective Multiple Test Procedure. In: Scandinavian Journal of Statistics, 6 (2), S. 65-70.
- Kada, Olivia (2010): Sex Role Orientation, Stressors and Resources in Male and Female Nurses. In: Review of Psychology, 17 (2), S. 97-104.
- Keller, Anita C./Meier, Laurenz L./Elfering, Achim/Semmer, Norbert K. (2019): Please Wait Until I am Done! Longitudinal Effects of Work Interruptions on Employee Well-Being. In: Work & Stress, 33 (1), S. 1-20. https://doi.org/10.1080/02678373.2019.1579266
- Li, Wen-Dong/Schaubroeck, John M./Xie, Jia Lin/Keller, Anita C. (2018): Is Being a Leader a Mixed Blessing? A Dual-Pathway Model Linking Leadership Role Occupancy to Well-Being. In: Journal of Organizational Behavior, 39 (8), S. 971-989. https://doi.org/10.1002/job.2273
- Mir, Eva/Kada, Olivia/Brunkel, Hella/Kohlmann, Elke/Kohlmann, Carl-Walter (2016): Wie nehmen Auszubildende der Altenpflege die Arbeits- und Organisationsstrukturen in der Praxis wahr? In: HeilberufeScience, 7 (2), S. 83-87. https://doi.org/10.1007/s16024-016-0263-x
- Molnar, Martina/Prinkel, Maria/Friesenbichler, Herbert (2013): Evaluierung psychischer Belastungen. Die Arbeits-Bewertungs-Skala ABS Gruppe. Wien: AUVA.
- Nielsen, Karina/Randall, Raymond John/Holten, Ann-Louise/González, Eusebio Rial (2010): Conducting Organizational-Level Occupational Health Interventions. What Works? In: Work & Stress, 24 (3), S. 234-259. https://doi.org/10.1080/02678373.2010.515393
- Prümper, Jochen/Hartmannsgruber, Klaus/Frese, Michael (1995): KFZA Kurz-Fragebogen zur Arbeitsanalyse. In: Zeitschrift für Arbeits- und Organisationspsychologie, 39 (3), S. 125-132.
- Prümper, Jochen/Schneeberg, Tom (2020): Validierung eines interaktiven Abstimmungssystems zur Diagnose psychischer Belastung mittels KFZA. In: Diagnostica, 66 (2), S. 172-177. https://doi. org/10.1026/0012-1924/a000245
- Prümper, Jochen/Vowinkel, Julia (2019): EVALOG Evaluierung psychischer Belastung im Dialog nach dem österreichischen ArbeitnehmerInnenschutzgesetz (ASchG) für Kleinstbetriebe. Wien: AUVA.
- Prümper, Jochen/Schneeberg, Tom/Prümper, Alina Marie (2021): EVALOG Beurteilung der Gebrauchstauglichkeit, Verständlich- und Handhabbarkeit, Teilnahmemotivation und emotionalen Reaktion bei der Evaluierung psychischer Belastung aus Sicht von Evaluierenden und Beschäftigten. In: Prävention und Gesundheitsförderung, 16 (4), S. 334-343. https://doi.org/10.1007/s11553-020-00817-6
- Prümper, Jochen/Strobach, Thomas/Vowinkel, Julia (2018): Motivation von KleinunternehmerInnen zur Teilnahme an der Entwicklung eines Verfahrens zur Evaluierung psychischer Belastung. In: Trimpop, Rüdiger/Kampe, Jana/Bald, Moritz/Selinger, Iris/Effenberger, Georg (Hg.): Psychologie der Arbeitssicherheit und Gesundheit. Asanger: Kröning, S. 489-492.
- Rieder, Kerstin/Kraus, Sylvia/Vogl, Gerlinde (2020): On the Road Again: Wie kann die Arbeitsgestaltung zur Arbeitsfreude bei mobiler Arbeit beitragen? In: Organisationsberatung, Supervision, Coaching, (2020) 27, S. 21-36. https://doi.org/10.1007/s11613-020-00635-x
- Rosenthal, Robert/Fode, Kermit L. (1963): The Effect of Experimenter Bias on the Performance of the Albino Rat. In: Behavioral Science, 8 (3), S. 183-189. https://doi.org/10.1002/bs.3830080302

Uhle, Thorsten/Treier, Michael (2019): Betriebliches Gesundheitsmanagement. Wiesbaden: Springer. https://doi.org/10.1007/978-3-658-25410-0

Wegge, Jürgen/Shemla, Meir/Haslam, Alexander (2014): Leader Behavior as a Determinant of Health at Work: Specification and Evidence of Five Key Pathways. In: German Journal of Research in Human Resource Management, 28 (1-2), S. 6-23. https://doi.org/10.1177/239700221402800102