

Unterstützendes Führungsverhalten schulischer Leitungskräfte für die Arbeit professioneller Lerngemeinschaften im Kollegium

Julia Warwas · Christoph Helm · Christian Schadt

Angenommen: 29. Januar 2019 / Online publiziert: 22. Februar 2019
© Der/die Autor(en) 2019

Zusammenfassung Ergänzend zu einer primär outcome-orientierten Erforschung von Professionellen Lerngemeinschaften wendet sich der Beitrag günstigen inner-schulischen Entwicklungsbedingungen für die Arbeit dieser Gemeinschaften zu. Dabei interessiert vor allem, welche prädiktive Kraft unterstützende Führungspraktiken der Schulleitung für variierende Ausprägungen konstitutiver Merkmale von Professionellen Lerngemeinschaften besitzen und worin diese Führungspraktiken konkret bestehen. Um diesen Fragen nachzugehen, werden im Theorieteil des Beitrags sowohl Kerndimensionen von Professionellen Lerngemeinschaften als auch mögliche Ansatzpunkte ihrer gezielten Förderung durch eine Synopse vorliegender Systematisierungen und Einzelstudien beschrieben. Die Literatursichtung lässt u. a. erkennen, dass sich eine unterstützende Führung nicht auf die Bereitstellung struktureller Ressourcen beschränkt, sondern zudem die Stärkung vertrauensvoller Beziehungen, inhaltliche Impulse für die Unterrichtsentwicklung sowie die Teilung von Führungsverantwortung beinhaltet. Die Ergebnisse von Mehrebenen-Strukturgleichungsmodellen ($N=395$ Lehrkräfte) dokumentieren erwartungskonform praktisch bedeutsame Zusammenhänge zwischen wahrgenommenem Führungsverhalten

Prof. Dr. J. Warwas · C. Schadt, M.Sc.
Wirtschaftswissenschaftliche Fakultät, Professur für Wirtschaftspädagogik mit dem Schwerpunkt Berufliches Lehren und Lernen, Georg-Augustus-Universität Göttingen, Platz der Göttinger Sieben 5, 37073 Göttingen, Deutschland

Prof. Dr. J. Warwas
E-Mail: julia.warwas@uni-goettingen.de

C. Schadt, M.Sc.
E-Mail: christian.schadt@uni-goettingen.de

Ass.-Prof. Dr. C. Helm (✉)
Abteilung für Bildungsforschung, Johannes Kepler Universität Linz, Altenberger Straße 69, 4040 Linz-Auhof, Österreich
E-Mail: christoph.helm@jku.at

und Kerndimensionen professioneller Gemeinschaften im Kollegium, auch wenn die Einschätzungen der Lehrkräfte die theoretisch angenommene Differenzierung von vier Facetten des Führungsverhaltens nicht exakt widerspiegeln.

Schlüsselwörter Professionelle Lerngemeinschaften · Bedingungsbezogene Forschung · Unterstützende Führung · Schulleitung

Principals' supportive leadership for the work of teachers' professional learning communities

Abstract Complementing prevailingly outcome-oriented research on professional learning communities (PLCs), this paper focuses on conditions within school organizations that promote the communities' work. Of prime interest is the predictive value of principals' supportive leadership for variations of constitutive PLC elements and, furthermore, the specific features that characterize supportive leadership. To answer these questions, the theoretical part of the paper describes core dimensions of PLCs and potential approaches of targeted support by reviewing available classifications and empirical investigations. Our review reveals, among other things, that supportive leadership is not restricted to providing structural resources but also includes establishing trustful relationships, contributing ideas to instructional improvements, and sharing responsibilities. As expected, results from multilevel structural equation models ($N=395$ teachers) document significant associations between perceived leadership behaviours and core dimensions of professional communities among faculty staff, even though staff members' ratings do not fully reflect the theoretically assumed distinction of four different facets of supportive leadership.

Keywords Professional Learning Communities · Research on Antecedents · Supportive Leadership · Principals

1 Hintergrund und Fragestellungen

Mit der Gewährung operativer Entscheidungs- und Gestaltungsfreiräume zur Erreichung verbindlicher Bildungsstandards sind Einzelschulen dazu aufgerufen, Maßnahmen der Organisations-, Personal- und Unterrichtsentwicklung zu ergreifen (zsf. Altrichter und Helm 2011). Als wichtige Motoren dieser Entwicklungsprozesse gelten *Professionelle Lerngemeinschaften (PLGen) von Lehrkräften*, die mit dem Ziel einer langfristigen Veränderung der Unterrichtspraxis eine systematische und reflektierte Kooperation pflegen (Berkemeyer et al. 2011; Fussangel und Gräsel 2014; Buhren 2015). Sowohl Meta-Analysen internationaler Forschungsbeiträge als auch erste Studien aus dem deutschsprachigen Raum bestärken dabei optimistische Erwartungen an die Arbeit von PLGen. Sie dokumentieren positive Zusammenhänge zwischen der Intensität von PLG-Aktivitäten der Lehrkräfte und dem Niveau erzielter Testleistungen von Schüler/innen (z. B. Lomos et al. 2011a) sowie der Qualität der Unterrichtsgestaltung (z. B. Warwas und Helm 2018; Webs und Holtappels 2018). Der naheliegenden Anschlussfrage, welche schulinternen *Voraussetzungen* intensive

PLG-Aktivitäten begünstigen, widmeten sich bis vor wenigen Jahren aber vorrangig pragmatisch akzentuierte Werke (z. B. Eaker et al. 2002; Hord und Sommers 2008) und deskriptive Einzelfallstudien (z. B. Hipp et al. 2008). So stellen etwa die vielzitierten Grundlagenwerke zu PLGen von DuFour und Eaker praktische Handlungsanleitungen für die Etablierung professioneller Lerngemeinschaften durch Lehrkräfte und Schulleiterinnen dar, die fast überwiegend auf narrativen Erfahrungsberichten der Autoren an wenigen Schulen fußen.

Trotz eines erkennbaren Zuwachses an bedingungsbezogener PLG-Forschung gelten insbesondere *unterstützende Führungspraktiken schulischer Leitungskräfte* nach wie vor als unterbelichtet (z. B. Vangrieken et al. 2017). Dass Untersuchungen, die sich explizit mit der Rolle und dem Handeln der Schulleitung gegenüber PLGen auseinandersetzen, eher selten anzutreffen sind, scheint insofern verwunderlich, als der internationale Bestand an Studien zur sog. Wirksamkeit der Schulleitung mittlerweile beachtlich ist (Leithwood und Sun 2012). Zu den zentralen und konvergenten Ergebnissen dieser Studien zählen *indirekte*, d. h. über Arbeitseinstellungen, -prozesse und -strukturen der Lehrkräfte vermittelte Effekte von Führungspraktiken auf die Zielgröße der Schülerleistungen. Erkennt man jedoch PLGen als fruchtbare Elemente der innerschulischen Zusammenarbeit von Lehrkräften an, die ihrerseits mit schülerseitigen Lernzuwächsen in überzufälliger Verbindung stehen (s. oben), und bedenkt man ferner, dass auch eine planvolle Personalentwicklung zum Aufgabebereich der Schulleitung gehört (Berkemeyer et al. 2015), muss die Rolle schulischer Führungskräfte für eine produktive Arbeit von Lehrkräften in PLGen stärker in den Fokus empirischer Forschung rücken.

Speziell die deutschsprachige Schulleitungsforschung ist aufgrund des bestehenden Mangels an thematisch einschlägigen Untersuchungen (siehe Wissinger 2014) gegenwärtig darauf angewiesen, Befunde zu wirkungsvollen Führungspraktiken aus anderen Bildungssystemen zu rezipieren, ohne sicher sein zu können, dass jene auf schulische Handlungskontexte in Deutschland, Österreich oder der Schweiz übertragbar sind (kritisch z. B. Klein 2016). Angesichts der hohen Verantwortung, welche die Bildungspolitik schulischen Führungskräften für den Anstoß und die Begleitung gelingender Schulentwicklung zuweist (Brauckmann 2012), scheint die Gewinnung wissenschaftlich geprüfter Empfehlungen für die Führungspraxis dringend geboten. Eine besondere Herausforderung besteht allerdings darin, dass in verfügbaren internationalen Studien sowohl konstitutive Merkmale von PLGen als auch fruchtbare Fördermöglichkeiten durch die Schulleitung teilweise unterschiedlich operationalisiert werden (kritisch z. B. Slegers et al. 2013; Vanblaere und Devos 2016).

Vor diesem Hintergrund geht der vorliegende Beitrag drei Fragen nach (1) Welche Kerndimensionen zur Beschreibung von PLG-Aktivitäten einerseits sowie von förderlichen Führungspraktiken andererseits kristallisieren sich als Schnittmenge (inter-)nationaler Forschungsbeiträge heraus? (2) Welche Dimensionen der Arbeit in PLGen sowie des unterstützenden Verhaltens von schulischen Leitungskräften lassen sich aus der Perspektive von Lehrkräften empirisch unterscheiden? (3) Stehen die Ausprägungen wahrgenommener Führungspraktiken in systematischem Zusammenhang mit der berichteten Arbeitsweise in PLGen?

Der Theorieteil des Beitrags setzt sich zunächst in einem synoptischen Zugriff mit Klassifikationsmerkmalen von PLGen (Abschn. 2.1) und Funktionsbeschreibungen einer unterstützenden Führung (Abschn. 2.2) auseinander. Die im empirischen Teil berichteten Ergebnisse stützen sich auf Daten von 395 Lehrkräften aus 47 berufsfachlich spezialisierten Abteilungen an beruflichen Schulen in Deutschland, die mithilfe von *doppelt latenten Mehrebenen-Strukturgleichungsmodellen* ausgewertet wurden (Abschn. 4.1 und 4.2). An die Diskussion der Ergebnisse schließen sich im letzten Kapitel Anregungen für künftige Untersuchungen und die schulische Führungspraxis.

2 Theoretische Basis

2.1 Kerndimensionen innerschulischer PLGen

Trotz stetig steigender Publikationszahlen mangelt es der PLG-Forschung bis dato an einer konsentierten Definition ihres Gegenstandsbereichs (hierzu kritisch z. B. Vanblaere und Devos 2016; Salleh et al. 2017). Eine erste Annäherung an das Konzept der PLG erlaubt die semantische Erschließung seiner begrifflichen Bestandteile im Sinne einer vorläufigen Nominaldefinition (ebd.). Demnach verbinden fortwährende, kollektive Erfahrungsgewinne (*Lernen*) die Mitglieder einer Gruppe, welche durch gemeinsames „Fühlen, Streben und Urteilen verbunden“ sind (*Gemeinschaft*) und als „Experten in einem umgrenzten Gebiet“ ihre Lernanstrengungen auf die Vervollkommnung domänenspezifischer Kenntnisse und Fähigkeiten konzentrieren (*professionell*) (Bonsen und Hübner 2012, S. 169). Im pädagogischen Handlungskontext verweist die Bezeichnung PLG somit auf „groups of teachers who share and critically question their practice in an on-going, reflecting, collaborative, learning-oriented way to promote their growth and skill“ (Rone 2009, S. 4). Offen bleibt jedoch, wo die Grenzlinien der Domäne verlaufen, auf die sich die Expertise der Gemeinschaft erstreckt (etwa ein spezielles Fach, ein zertifizierter Bildungsgang oder eine Schulungsform). Diese Unschärfe bedingt eine große Vielfalt an Untersuchungsobjekten in der PLG-Literatur, angefangen von kleinen Projektteams ohne formalorganisatorische Verankerung über Kollegien ganzer Schulen bis hin zu schulübergreifenden Netzwerken (Vangrieken et al. 2017). Im deutschsprachigen Schulkontext richtet sich die Aufmerksamkeit überwiegend auf fachwissenschaftlich und fachdidaktisch spezialisierte Lehrkräfteteams, die an großen Schulen abteilungsförmig organisiert sind (Bonsen und Hübner 2012; Buhen 2015). An Berufsschulen und Berufsschulzentren, die der vorliegende Beitrag beleuchtet, handelt es sich hierbei um Abteilungen, deren Mitglieder die pädagogisch-didaktische Gestaltung von Ausbildungsgängen in einer bestimmten Domäne des beruflichen Bildungssystems verantworten (bspw. in der kaufmännisch-verwaltenden Domäne oder gewerblich-technischen Domäne; siehe hierzu Röder 2017; Tenberg 2017).

Mithilfe von merkmalsgestützten, operationalen Definitionen haben zahlreiche Autor/innen die Unzulänglichkeiten einer rein semantischen Erschließung des PLG-Konzepts für empirische Analyse Zwecke zu überwinden gesucht (im Überblick Warwas und Helm 2018). Zu den häufig zitierten Vorschlägen zählen die beiden

nur teilweise kongruenten *Fünf-Komponenten-Modelle* von Hord (1997) sowie von Kruse et al. (1995). Nach Auffassung von Hord bestehen diese Komponenten aus (1) *geteilten Werten und Visionen* hinsichtlich Lehren und Lernen im Kollegenkreis, (2) *unterstützenden organisatorischen Bedingungen* für eine (3) *enge Kooperation* und (4) *gemeinsames forschungsbasiertes Lernen* der Lehrkräfte sowie (5) einer *unterstützenden, geteilten Führung* im Sinne dezentralisierter Führungsverantwortung im Zuständigkeitsbereich der PLG. Kruse et al. (1995) sehen neben (1) *geteilten Normen und Werten* der PLG-Mitglieder hingegen (2) einen *reflektierenden Dialog* über Unterricht, (3) die *Deprivatisierung der pädagogischen Praxis* als Öffnung des eigenen Unterrichts für kollegiales Feedback, (4) eine *gemeinsame Fokussierung auf eine Förderung des Schülerlernens* sowie (5) eine *kooperative Arbeitsweise* als wesentliche Komponenten vor. In der Literatur finden sich darüber hinaus eklektische Ansätze, welche die gerade skizzierten Modelle verschmelzen (z. B. Jackson und Tasker 2002) oder um diverse zusätzliche Komponenten anreichern (z. B. Stoll et al. 2006; Mitchell und Sackney 2007; Sigurdardottir 2010).

Mit der Zielsetzung, kategoriale Schnittmengen zwischen den heterogenen Systematisierungsvorschlägen zu ermitteln und dabei von eher randständigen Merkmalen zu abstrahieren, lassen sich inhaltsanalytisch drei Kerndimensionen von PLGen eingrenzen (Achinstein 2002; Lavié 2006; Zheng et al. 2016; Warwas und Helm 2018 und dortige Quellennachweise):

- Die *aktionale Dimension, Kooperative Entwicklung*, umfasst alle gemeinschaftlichen Anstrengungen, welche der steten Verbesserung schülerseitiger Lernprozesse wie auch der Erweiterung von hierfür erforderlichen professionellen Kompetenzen dienen. Diese Dimension beinhaltet folglich zum einen die gemeinsame Planung und Implementierung von Unterrichtsreihen und Lehrstrategien. Zum anderen erstreckt sie sich auf Aktivitäten, die im Sinne einer „continuous collective inquiry“ (Hord 1997) der unablässigen, kritischen Revision unterrichtlicher Praxis und angestoßener Veränderungen derselben dienen. Hierzu gehören neben einem reflexiven, dialogischen Austausch auch gegenseitige Hospitationen, kollegiale Fallberatung oder kriteriengestützte Evaluationsmaßnahmen.
- Die *ideelle Dimension, Normativer Konsens*, verweist auf *gruppenintern* geteilte pädagogische Überzeugungen und Werte sowie auf konvergierende Vorstellungen über die mit der Zusammenarbeit verfolgten Ziele. Obwohl diese Dimension offenkundige Schnittmengen mit dem theoretischen Konstrukt der Schulkultur aufweist, ist aber nicht von einer Deckungsgleichheit auszugehen.¹ Die ideelle Di-

¹ Eine Schulkultur ist zunächst als gesamtschulisches Merkmal auf einer anderen Analyseebene angesiedelt (van Houtte 2005). Schulkulturen gründen im Sinne organisationstypischer Orientierungspunkte von Handlungsvollzügen in den Werten und Überzeugungen sämtlicher oder zumindest der meisten Stakeholder der betrachteten Organisation (also etwa auch der Schüler(innen) und Eltern; vgl. Carpenter 2015). Weiterhin bieten Schulkulturen definitionsgemäß *multiple* Orientierungspunkte für die Alltagspraxis, da sich die geteilten normativen Vorstellungen auf unterrichtliche, schulorganisatorische und sogar außerschulische Handlungsvollzüge beziehen (van Ackeren et al. 2008, S. 342). Die im engeren Sinne auf pädagogische Handlungsziele und Leistungserwartungen zugeschnittene *Erziehungs- bzw. Lehr-Lern-Kultur* stellt demnach nur eine von mehreren Facetten einer Schulkultur dar (ebd.). Der intuitiven Annahme der Homogenität einer Schulkultur widersprechen indes Befunde der Fachkulturforschung. Sie dokumentieren vielfältige fach(bereichs-)gebundene Subkulturen innerhalb schulischer Organisationen mit stark divergierenden, primär fachintern konsistenten Überzeugungs- und Handlungsmustern (Lüders 2007).

mension einer PLG beschränkt sich stattdessen auf Leitideen des von den PLG-Mitgliedern gemeinsam verantworteten Arbeitsbereichs, das heißt bspw. im Falle fachbereichsgebundener PLGen auf Leitideen der Gestaltung und Verbesserung von Lehr-Lern-Prozessen in der unterrichteten Domäne unter dem Globalziel der „Förderung des Schülerlernens“. PLGen mit starkem normativen Konsens verfügen damit über eine starke eigenständige Lehr-Lern-Kultur, die mit der weiter gefassten, schulweiten Lehr-Lern-Kultur (soweit als solche vorhanden) kompatibel sein kann, aber nicht zwingend muss (van Ackeren et al. 2008).

- Die strukturelle Dimension, *Tragfähige Infrastruktur*, beschreibt das operative Fundament einer PLG. Sie beinhaltet Stützfaktoren der zielgerichteten Durchführung kooperativer Arbeitsprozesse, so etwa klare Regeln und Zuständigkeiten, transparente Entscheidungsfindungen und schnelle Informationsflüsse, effiziente Moderation von Teamsitzungen oder eine reibungslose Projektkoordination.

Im Ergebnis zeigt der Abgleich operationaler Definitionen, dass die identifizierten Kerndimensionen die rein semantischen Konnotationen einer PLG sowohl differenzierter abbilden als auch über diese hinausreichen. Während die *kollektiven Erfahrungsgewinne* aus Maßnahmen der Unterrichtsentwicklung erwachsen, die nicht nur gemeinschaftlich geplant und konsistent umgesetzt, sondern auch kontinuierlich überprüft und ggf. revidiert werden, konzentriert sich das *gemeinsame Streben* auf die Optimierung schülerseitiger Lernprozesse und konkretisiert sich in den von der Gruppe verfolgten Arbeitszielen. Die in der Nominaldefinition nicht abgebildete (*infra*-)strukturelle Dimension berücksichtigt zusätzlich Facetten einer effizienten Arbeitsweise mit ausreichenden Ressourcen (Eaker et al. 2002; Cormier und Olivier 2009).

Auf Basis der Kerndimensionen lassen sich die einleitend beschriebenen empirischen Hinweise auf unterrichtliche Qualitäts- und schülerseitige Leistungssteigerungen besser nachvollziehen. Obwohl die zugrundeliegenden Mechanismen noch nicht in Form von Mediatoranalysen untersucht wurden, dürften Qualitäts- und Leistungssteigerungen u. a. auf nachweisbare individuelle Zuwächse an pädagogischem und (fach-)didaktischem Wissen der PLG-Mitglieder zurückzuführen sein, welche sich infolge intensiven Austauschs und gegenseitiger Unterstützung einstellen (Bolam et al. 2005; Bonsen und Rolff 2006; Valckx et al. 2018). Ferner ist auf die Systematik, Dauer und Wirkungsbreite von Unterrichtsentwicklungsprozessen zu verweisen, die im Unterschied zu punktuellen Initiativen einzelner Lehrkräfte flächendeckend für einen ganzen Fachbereich oder Bildungsgang konzipiert, realisiert und stetig optimiert werden (Bonsen und Hübner 2012).

Deskriptive Befunde zum Verbreitungsgrad von PLGen in der schulischen Praxis zeigen jedoch, dass nur wenige Lehrkräftegruppen die Erkennungsmerkmale von PLGen in idealtypischer Ausprägung aufweisen (z. B. Lomos et al. 2011b; Warwas und Helm 2018). In Anbetracht ihres begründet erwartbaren Nutzens stellt sich daher die Frage, welche schulinternen Bedingungen die Entstehung von PLGen überzufällig begünstigen bzw. erschweren. Möchte man dabei auch Ansatzpunkte für einzelschulische Interventionen identifizieren, lässt sich konkreter nach Führungsstrategien fragen, mit denen die für die Qualitätsentwicklung gesamtverantwortliche Schulleitung die Lehrerverkooperation gezielt intensivieren kann (z. B. Dubs

2009; Wissinger 2014). Folgt man der Auffassung von Cormier und Olivier (2009, S. 4), dass das Handeln des schulischen Leitungspersonals zu den „major theoretical constructs associated with the implementation of a (...) professional learning community“ gehört, müssen die zu spezifizierenden Führungsstrategien entsprechend der oben geschilderten Kerndimensionen von PLGen grundsätzlich dazu geeignet sein, in diesen Gemeinschaften die systematische, kooperative Entwicklung von Unterricht und Lehrendenkompetenzen zu fördern, den normativen Konsens zu stärken und eine tragfähige Infrastruktur auf- bzw. auszubauen.

2.2 Ansatzpunkte unterstützenden Führungsverhaltens

In Beiträgen zu unterstützender Führung von PLGen lassen sich neben meist bildhaften Beschreibungen multipler Rollen der Schulleitung (z. B. Hord und Sommers 2008; Cormier und Olivier 2009) diverse Einzelstudien ausmachen, die verhaltensbezogene Führungsstrategien diskutieren und indikatorgestützt untersuchen. Hierbei wählen sie jedoch entweder unterschiedliche Strategien aus oder verwenden trotz inhaltlicher Verwandtschaft der ausgewählten Strategien abweichende Bezeichnungen (hierzu kritisch z. B. Vanblaere und Devos 2016). Die folgenden vier Ansatzpunkte unterstützenden Führungsverhaltens stellen eine Synopse häufig thematisierter Führungsstrategien dar. Sie erlaubt die Einordnung von Führungsstrategien und zugehöriger empirischer Befunde in Kategorien, die nach *funktional unterscheidbaren Handlungsoptionen* zur Unterstützung von PLGen in schulischen Handlungskontexten gegliedert sind.

- *Strukturelle Erleichterungen für die operativen Arbeitsprozesse der PLG*: Sollen gemeinschaftliche Anstrengungen zur Verbesserung der Unterrichtsqualität planvoll, regelmäßig und langfristig betrieben werden, muss die Schulleitung ausreichende Gelegenheiten hierfür schaffen (Olivier et al. 2010). Dies erfolgt insbes. durch die Ausstattung der PLGen sowie Entlastungen der PLG-Mitglieder von administrativen Aufgaben (Buhren 2015). Grundsätzlich fallen derlei Maßnahmen somit in den Gestaltungsbereich der *arbeitsorganisatorischen Voraussetzungen* (z. B. Berkemeyer et al. 2015; Gray et al. 2016). Aus der Forschung zum unterstützenden Führungsverhalten ließe sich hier der Ansatz der *structural leadership* einordnen (vgl. Bryk et al. 1999). Dieser beschreibt eine Führungspraxis, die im Sinne eines ökonomischen Managements kooperatives Arbeiten durch die Bereitstellung zeitlicher, materieller und räumlicher Ressourcen grundsätzlich ermöglicht (ebd.). Diese Auffassung findet sich in internationalen (z. B. Barton und Stepanek 2012) wie auch deutschsprachigen Publikationen zu PLGen (z. B. Bensen und Hübner 2012). Eine arbeitsorganisatorisch unterstützende Führung erwies sich sowohl in qualitativen (Wells und Feun 2007) als auch quantitativen (Bolam et al. 2005) Studien als eng mit der Implementierung und dauerhaften Arbeit von PLGen verbunden. So berichten etwa Zhang et al. (2017) auf Basis einer Interviewstudie davon, dass Lehrkräfte als eine der größten Hürden für gelingendes Arbeiten in PLGen die fehlende Bereitstellung struktureller Ressourcen (v. a. Zeit zur Zusammenarbeit) durch die Schulleitung nennen.

- *Impulse für die Unterrichtsentwicklung:* Um Aktivitäten der PLGen auf das pädagogische Kerngeschäft zu konzentrieren, können Leitungskräfte *inhaltliche Anstöße* geben (z. B. Supovitz et al. 2009; Vanblaere und Devos 2016; Valckx et al. 2018). Ohne dabei in didaktischen Detailfragen zu bevormunden, können Schulleitende bspw. den reflektierten Dialog über guten Unterricht und die bestmögliche Förderung der Schüler/innen durch Fragen und Ideen anregen sowie auf Basis von Unterrichtsbesuchen konstruktive Rückmeldungen zu errungenen Erfolgen der Entwicklungsarbeit und verbleibenden Verbesserungspotentialen geben (z. B. Bryk et al. 1999; Thompson et al. 2004; Louis et al. 2010). Dementsprechend thematisieren einschlägige Studien meist unter der Bezeichnung *instructional leadership* (z. B. Vanblaere und Devos 2016, 2018; Zheng et al. 2016; Valckx et al. 2018) ein Führungsverhalten, bei dem Führungskräfte mit den Lehrkräften wichtige Themen der Unterrichtsarbeit diskutieren sowie auf eine Weiterentwicklung des Curriculums und der Lehrendenkompetenzen hinwirken, um Unterrichtsprozesse zu verbessern (Vanblaere und Devos 2018). Zugehörige Befunde verweisen darauf, dass sich sachinhaltliche Führungsimpulse positiv auf die zielorientierte Arbeit in PLGen auswirken kann, so etwa auf die Häufigkeit unterrichtsbezogener Kooperation und die gemeinsame Reflexion des Unterrichtshandelns (Zheng et al. 2016; Vanblaere und Devos 2018). Ebenso kommen Mitchell und Sackney (2016) sowie Zheng et al. (2016) in qualitativen Erhebungen an verschiedenen Schularten zu dem Befund, dass Leitungspersonen durch eine Betonung von Themen, die für das Lehren und Lernen relevant sind, sowohl zur Implementierung als auch zur Aufrechterhaltung von gut funktionierenden PLGen beitragen können. Damit helfen sie Lehrkräften dabei, den Fokus auf das Ziel der Verbesserung des Unterrichts und der Schaffung von Lerngelegenheit für Schüler/innen aufrechtzuhalten.
- *Vertrauens- und Beziehungsstiftung:* Selbst ergiebige inhaltliche Anregungen dürften für die pädagogische Praxis folgenlos bleiben, wenn das professionelle Verhältnis der Lehrkräfte von Misstrauen oder Unsicherheit geprägt ist (Wells und Feun 2007; Pröbstel und Soltau 2012; Wang 2015). Substanzielle Auseinandersetzungen mit unterrichtlichen Entwicklungsbedarfen und das erfolgreiche Ringen um problemadäquate Lösungsstrategien verlangen nach Offenheit, lassen argumentative Kontroversen zu und können die professionelle Autonomie und Identität einzelner Mitglieder zeitweilig erschüttern (Cranston 2011; Gray et al. 2016). Sie setzen daher vertrauensbasierte interpersonelle Beziehungen voraus, welche die Gruppenkohäsion auch in schwierigen Arbeitsphasen gewährleisten (Hallam et al. 2015; Zheng et al. 2016). Aus der Einsicht in diese *arbeitsklimatischen Grundlagen* tiefgreifender, kontinuierlicher Lernprozesse im Kollegium (Berkemeyer et al. 2015) erwächst die Aufgabe der Schulleitung, auch die sog. relationalen Ressourcen der Lerngemeinschaft zu stärken (z. B. Wahlstrom und Louis 2008; Li et al. 2016). Wenn das Führungshandeln darauf abzielt, die Beziehungsqualität unter Kolleg/innen zu verbessern, beeinflusst es die Arbeit von PLGen indirekt über die Stärkung einer wichtigen Facette der Schulkultur, der sog. Beziehungskultur (van Ackeren 2008). Zu diesem Zweck sollte das eigene Kommunikations- und Kooperationsverhalten der Schulleitung gegenüber Lehrkräften zu respektvoller, fehlerfreundlicher und lösungsorientierter Zusammenarbeit anregen (z. B. Thompson et al. 2004; Supovitz et al. 2009) und von

persönlicher Offenheit, Lern- und Experimentierfreude zeugen (z. B. Moller und Pankake 2006; Cranston 2011). Um Vertrauen zu stiften, sollte das Auftreten der Schulleitung ferner mit proklamierten Werten und den Leitbildern der Schule übereinstimmen (z. B. Thompson et al. 2004; Cranston 2011; Vanblaere und Devos 2016). Bei Konflikten muss sie zudem moderierend eingreifen (z. B. Scribner et al. 2002; Moller und Pankake 2006). Die orientierende und motivierende Funktion eines modellhaften Vorlebens von gemeinschaftsförderlichen Werten und Praktiken wird vielfach als *transformational leadership* bezeichnet (Vanblaere und Devos 2018; Zheng et al. 2016; Valckx et al. 2018). Dass sich ein derartiges Führungsverhalten positiv auf wichtige Merkmale der Arbeit von PLGen auswirken kann, legen Studien nahe, in denen bspw. die Intensität der Zusammenarbeit oder das Ausmaß der Öffnung des eigenen Unterrichts gegenüber den Kolleginnen und Kollegen analysiert wurden (Zheng et al. 2016; Vanblaere und Devos 2018). Valckx et al. (2018) kommen zu dem Ergebnis, dass sich Schulleitungen von sog. „high-perceived PLC departments“ (2018, S. 47) durch vergleichsweise starke *transformational leadership* auszeichnen. Hierdurch gelingt es ihnen besser als anderen Leitungskräften, kooperative Arbeitsweisen wie etwa einen reflektierten Dialog zwischen den Lehrkräften anzuregen. Die Stärkung eines innerschulischen Gemeinschafts- und Zugehörigkeitsgefühls durch ein Arbeitsumfeld, welches Lehrkräfte Sicherheit, Verlässlichkeit und Geborgenheit erleben lässt, findet sich auch im Ansatz der *caring leadership* (Louis et al. 2016). Sie bietet Lehrkräften annahmegemäß einen geschützten Raum, in dem die Öffnung eigenen Unterrichts und ein angst- und vorwurfsfreies Hinterfragen unterrichtlicher Praxis erleichtert wird, auch wenn entsprechende Zusammenhänge speziell für die Arbeit von PLGen empirisch noch nicht überprüft wurden.

- *Partizipative Führung*: Um nachhaltige Veränderungen des pädagogischen Kerngeschäfts zu ermöglichen, müssen die in den PLGen generierten Entwicklungsideen frühzeitig gehört, im Entwicklungskonzept der Schule ernsthaft berücksichtigt und im Fachbereich konsequent umgesetzt werden. Hierzu muss die Schulleitung bereit sein, Führungsverantwortung zu teilen und Lehrkräfte mit *Mitwirkungsrechten* auszustatten (z. B. McLaughlin und Talbert 2007; Wahlstrom und Louis 2008; Louis et al. 2010). Ohne ihre Gesamtverantwortung für die Koordination und Kontrolle laufender Schulentwicklungsprozesse aufgeben zu müssen (Dubs 2009), sollten Führungskräfte PLG-Mitglieder als Expert/innen in Fragen des Lehrens und Lernens konsultieren und ihnen Entscheidungs- und Handlungsfreiräume zugestehen (z. B. Scribner et al. 2002; Fleming 2004). Diese Überlegungen finden sich im Konzept der *shared leadership* wieder, die in zahlreichen Publikationen zu unterstützenden Führungspraktiken für PLGen diskutiert wird (z. B. Bolam et al. 2005; Borchers 2009; Wahlstrom und Louis 2008). Tatsächlich gehört die Übertragung von Verantwortungsbereichen von der Schulleitung auf die PLG-Mitglieder zu den frühesten Forderungen der PLG-Forschung (Hord 1997). Hierdurch soll es Lehrkräften ermöglicht werden, zur Zielerreichung in Schulentwicklungsprozessen eigenaktiv, weitgehend selbstgesteuert und möglichst restriktionsfrei beizutragen. Die bereits angesprochenen Studien von Vanblaere und Devos (2018) sowie Valckx et al. (2018) – erstgenannte Autoren nutzen die Begriffe *transformational* und *shared leadership* synonym – liefern konsistente Hinweise darauf, dass

geteilte Führungsverantwortung positiv mit Merkmalen der Zusammenarbeit von Lehrkräften assoziiert ist, etwa mit der Beteiligungsintensität an reflektierender Kommunikation oder der Ausprägung eines Verantwortlichkeitsempfindens für die gemeinsame Zielerreichung. Ferner kommen auch Berkemeyer et al. (2015) auf Basis einer wissenschaftlichen Begleitstudie sowie Borchers (2009) mit einem Mixed-Methods-Ansatz zu dem Ergebnis, dass die Schulleitung den PLGen Entscheidungsspielräume gewähren und beratend zur Seite stehen sollte.

Die obige Synopse legt nahe, dass sich Führungsstrategien zur Förderung der Arbeit in PLGen *funktional* unterscheiden lassen. Auf einer *konzeptionellen* Ebene erscheint es demnach sinnvoll, Führungspraktiken danach zu klassifizieren, ob sie den gemeinschaftlichen Aktivitäten der PLG-Mitglieder *arbeitsorganisatorische* Unterstützung bieten (strukturelle Erleichterungen) oder *sachinhaltliche* Anregungen liefern (Impulse für die Unterrichtsentwicklung), ob sie die *arbeitsklimatischen* Bedingungen der Zusammenarbeit zu verbessern helfen (Vertrauens- und Beziehungsstiftung) oder aber die innerschulischen *Mitwirkungsrechte* der PLGen ausweiten (partizipative Führung).

Im Folgenden wird der Frage nachgegangen, ob sich diese Führungspraktiken auch *empirisch* differenziert abbilden lassen und ob die *datengestützt* unterscheidbaren Dimensionen unterstützender Führung einen statistisch bedeutsamen Effekt auf die Arbeit von PLGen haben. Die oben referenzierten Studien lassen hierüber nur begrenzt Aussagen zu, da sie jeweils nur einzelne oder wenige Führungsstrategien aus der oben vorgeschlagenen Systematik untersuchen und die hiermit verbundenen Verhaltensweisen jeweils unterschiedlichen, manchmal auch nur einer globalen, Sammelkategorie(n) zuordnen.

3 Ziele und Methodik der Untersuchung

3.1 Annahmen und Erkenntnisinteressen

Die theoretischen Vorüberlegungen lassen hohe Ausprägungen konstitutiver Merkmale von PLGen vorrangig in solchen Organisationseinheiten erwarten, in denen Schulleiter/innen die Zusammenarbeit der Lehrpersonen strukturell erleichtern, inhaltlich anregen, in ein vertrauensvolles, integrierendes Klima einbetten und Führungsverantwortung teilen. Die nachfolgenden Analysen dienen daher zunächst der Überprüfung der Dimensionsstruktur von PLGen und Führungsstrategien aus Lehrersicht, bevor die Erklärungsbeiträge unterstützender Führung für variierende Ausprägungen von PLG-Dimensionen ermittelt werden.

3.2 Stichprobe und Erhebungsinstrumente

Die Untersuchung wurde in Bayern/Deutschland im Sommer 2011 durchgeführt. Aus ökonomischen Gründen wurden *cluster samples* (Klumpenstichprobe, Bortz und Schuster 2010, S. 81) erfasst. Insgesamt nahmen 395 Lehrpersonen (45,5% weiblich; durchschnittlich 17 Jahre Unterrichtserfahrung, SD= 10,7 Jahre) an einer

Tab. 1 Beispielitems, deskriptive Statistiken und Reliabilitäten für die PLG-Dimensionen

	#	Beispielitem	M	SD	ω	ICC(1)	ICC(2)
KE	7	Wir bereiten Unterrichtseinheiten häufig gemeinsam vor	3,39	0,86	0,81	0,26	0,79
NK	5	Die didaktischen Vorstellungen im Fachkollegium lassen sich kaum auf einen gemeinsamen Nenner bringen (invers)	3,63	0,77	0,75	0,16	0,62
TI	9	Unsere Arbeitsabläufe sind gut strukturiert	4,14	0,76	0,87	0,31	0,75

Anzahl der Items je Skala, *M* Mittelwert, *SD* Standardabweichung, ω Omega, *KE* Kooperative Entwicklung, *NK* Normativer Konsens, *TI* Tragfähige Infrastruktur

Antwortskala: 1 = trifft gar nicht zu – 6 = trifft sehr zu

ICC(1) = Intraklassenkorrelation, ICC(2) = Reliabilität der aggregierten Individualurteile

schriftlichen Befragung teil. Sie stammen aus 47 beruflichen Fachbereichsgruppen (technisch-industriell: 11, sozial: 10, kaufmännisch: 20, technologisch: 6), die 34 berufsbildenden Schulen zuzuordnen sind, wobei in 23 Schulen ein Fachbereich, in 9 Schulen zwei Fachbereiche und in 2 Schulen drei Fachbereiche erfasst wurden. Im Mittel füllten je Fachbereich 8 Lehrkräfte (min. 4, max. 25) die Fragebögen aus. Nähere Informationen zu Stichprobenziehung, Repräsentativität und Rücklaufquote sind dem Anhang A1 zu entnehmen.

3.2.1 Erfassung der Arbeit in PLGen

Kerndimensionen professioneller Lerngemeinschaften wurden mit drei Subskalen erhoben (Tab. 1). Die zugehörigen Items wurden aus Studien zur Schulqualität und Lehrerverbundenheit in Deutschland adaptiert (Ditton o.J.; Steinert et al. 2003), da sich in der PLG-Forschung kein Instrument etabliert hat, das sowohl studienübergreifend einheitlich eingesetzt wird als auch psychometrisch validiert ist (Vangrieken et al. 2015).

3.2.2 Erfassung des Führungsverhaltens

Zur Erfassung der in Abschn. 2.2 beschriebenen Führungsstrategien wurden vier Subskalen verwendet, die sich auch zu einem Globalfaktor unterstützender Führung (UF) zusammenfassen lassen (siehe Tab. 2). Dabei handelt es sich um Eigenentwicklungen auf Basis der vorgenommenen Literatursichtung.

Alle Fragebogenitems sind im Anhang A1 dargestellt. Darüber hinaus befinden sich Korrelationstabellen der latenten Konstrukte in Anhang A2.

3.2.3 Analyseverfahren

Um die empirische Dimensionsstruktur der professionellen Zusammenarbeit sowie des Führungsverhaltens zu prüfen, wurden Zweiebenen-Strukturgleichungsmodelle geschätzt. Dabei kamen sog. *Doubly Latent-Modelle* zum Einsatz, um für den Mess- und Stichprobenfehler zu kontrollieren (Lüdtke et al. 2008). Darüber hinaus wurden Itemparcels gebildet und die konfigurale Messinvarianz über die Fachbereiche

Tab. 2 Beispielitems, deskriptive Statistiken und Reliabilitäten für die Führungsdimensionen

	#	Beispielitem	M	SD	ω	ICC(1)	ICC(2)
SE	3	Die Schulleitung versucht, die Lehrkräfte von allen nicht unmittelbar unterrichtsbezogenen Aufgaben zu entlasten	3,76	0,97	0,81	0,10	0,48
IU	3	Die Schulleitung diskutiert regelmäßig im Fachkollegium Fragen der Unterrichtsgestaltung	3,29	1,03	0,77	0,11	0,50
VB	5	Die Schulleitung schafft entspannte Arbeitssituationen, die auf gegenseitigem Vertrauen basieren	4,39	0,87	0,88	0,14	0,58
PF	3	Die Schulleitung bindet die Lehrkräfte in verantwortungsvolle Führungsaufgaben ein	4,20	0,97	0,82	0,06	0,35
UF	14	Globalfaktor zum Führungsverhalten mit allen Items	3,98	0,80	0,90	0,10	0,48

Anzahl der Items je Skala, *M* Mittelwert, *SD* Standardabweichung, ω Omega, *SE* Strukturelle Erleichterungen, *IU* Impulse für Unterrichtsentwicklung, *VB* Vertrauens- und Beziehungsstiftung, *PF* Partizipative Führung, *UF* Globalfaktor „Unterstützende Führung“
 Antwortskala: 1 = trifft gar nicht zu – 6 = trifft sehr zu
 ICC(1) = Intraklassenkorrelation, ICC(2) = Reliabilität der aggregierten Individualurteile

hinweg spezifiziert. Die Gründe hierfür sind in Anhang A3 ausführlich dargestellt. Die Schulebene wurde nicht zusätzlich berücksichtigt, da nur in rund einem Viertel der Schulen mehr als ein Fachbereich erfasst wurde. Damit bilden Fachbereiche und Schulen de facto meist dieselben Cluster.²

Da es sich bei den eingesetzten Skalen nicht um erprobte Messinstrumente handelt, wurde die Faktorenstruktur durch eine schrittweise Reduktion der Anzahl der hinter den Items vermuteten latenten Konstrukte geprüft. Die detaillierte Darstellung dieses Vorgehens findet sich in Anhang A4.

Um die Effekte des wahrgenommenen Führungsverhaltens auf die professionelle Zusammenarbeit in den Fachbereichen zu prüfen, wurden Zweiebenen-Strukturgleichungsmodelle spezifiziert, welche die Lehrer- und Fachbereichsebene modellieren. Da wir an der Vorhersage der professionellen Zusammenarbeit in Fachbereichen interessiert sind, stellt die Fachbereichsebene die im Folgenden fokussierte Analyseebene dar.

Alle Analysen wurden in Mplus 8 (Muthén und Muthén 1998–2017) unter Verwendung des ML-Schätzers ausgeführt. Der Anteil an fehlenden Werten liegt bei etwa 5%. Um die fehlenden Werte zu berücksichtigen, wurde die Parameterschät-

² Die Aufteilung der Varianz des Globalfaktors *unterstützendes Führungsverhalten* auf die drei Ebenen ergibt folgende Verteilung: 14% Schulebene, 11% Fachbereichsebene, 75% Lehrerebene. Dies entspricht insofern den Erwartungen, als dass die aus Lehrersicht eingeschätzte Unterstützung der Schulleitung den Fachbereichen gegenüber innerhalb einer Schule grundsätzlich ähnlich ausfallen sollte (14%). Gleichzeitig ist aber anzunehmen, dass Schulleitungen zwischen Fachbereichen differenzieren und diese zumindest graduell unterschiedlich stark unterstützen, sodass auch Lehrpersonen innerhalb eines Fachbereiches dieses Verhalten ähnlicher einschätzen (11%) als Lehrpersonen aus verschiedenen Fachbereichen derselben Schule.

zung unter der Missing at Random-Annahme mittels Full Information Maximum Likelihood (FIML)-Option durchgeführt.

Zur Beurteilung des Modellfits werden übliche Cut Off-Kriterien herangezogen (Hu und Bentler 1999; Little 2013): der χ^2/df -Wert ($\chi^2/df < 3$), der Bentlers Comparative Fit Index ($CFI \geq 0,90$), der Tucker Lewis Index ($TLI \geq 0,90$) und der Root Mean Square Error of Approximation ($RMSEA \leq 0,08$) sowie der Standardized Root Mean Square Residual auf Lehrerebene (Level 1, L1) und auf Fachbereichsebene (Level 2, L2) ($SRMR_{L1/L2} \leq 0,10$). Zur Interpretation der Reliabilität der verwendeten Skalen wird auf Lehrerebene das Omega (ω) (Geldhof et al. 2014) und auf Fachbereichsebene der ICC(2) (Lüdtke et al. 2006) herangezogen. Für beide Indizes gilt die Daumenregel, dass Werte über 0,70 auf eine akzeptable Reliabilität hindeuten (bspw. Lüdtke et al. 2006, S. 88). Grundsätzlich ist hervorzuheben, dass die Reliabilität (so auch die sehr niedrigen ICC(2)-Werte) der Skalen in den hier verwendeten *Doubly-Latent-Modellen* durch die Spezifikation des Mess- und Stichprobenfehlers mitmodelliert und hierdurch statistisch kontrolliert wird (Marsh et al. 2009, S. 776). Es ist allerdings darauf hinzuweisen, dass die Korrektur des Stichprobenfehlers in Fällen, in denen ein Sampling Ratio nahe 1 vorliegt, d.h. nahezu alle Fälle eines Custers (hier: alle Lehrkräfte eines Fachbereichs) befragt wurden, zu einer Überschätzung der Cluster-Ebeneneffekte führen kann (Lüdtke et al. 2008). Da für die vorliegende Stichprobe das Sampling Ratio nicht bekannt ist (siehe Anhang A1), ist nicht bestimmbar, ob die Korrektur des Stichprobenfehlers adäquat erfolgt oder die Gefahr überschätzter Effekte birgt. Die niedrigen ICC(2)-Werte sind allerdings insofern problematisch, als dadurch die Validität der Führungsskalen beeinträchtigt sein könnte. Aus diesem Grund haben wir die von Lüdtke et al. (2006) empfohlene Vorgehensweise angewandt und Zusatzanalysen durchgeführt, die zeigen, dass sich die untersuchten Effekte nicht wesentlich ändern, wenn man jene Cluster entfernt, die eine zu geringe Urteilerübereinstimmung (gemessen am $R_{WG(j)}$ und $AD_{M(j)}$) aufweisen. Die Ergebnisse der Zusatzanalysen sind in Anhang A5 dargestellt.

Da die Stichprobe auf Fachbereichsebene ($N=47$) sowie die mittlere Fachbereichsgröße ($N=8$) und die ICC(1)-Werte für die Führungsskalen relativ klein sind, wurden zudem Simulationsstudien durchgeführt (Anhang A6), die der Frage nachgehen, inwiefern die untersuchten Effekte (Bias-Analyse) und Signifikanzen (Power-Analyse) aufgrund dieser eher ungünstigen Stichprobenmerkmale verzerrt sind. Die Befunde zeigen, dass (1) der Bias vernachlässigbar ist und (2) die Signifikanzwerte lediglich für die Analysen mit der Skala „Kooperative Entwicklung“ aufgrund der niedrigen Power vorsichtig zu interpretieren sind. Daher ist begründet davon auszugehen, dass die im Folgenden berichteten Befunde nicht bzw. kaum durch die angeführten Stichprobenmerkmale verzerrt sind.

4 Empirische Befunde

4.1 Dimensionale Struktur der Zusammenarbeit und des Führungsverhaltens aus Sicht der PLG-Mitglieder

PLG-Kerndimensionen: Tab. 3 weist für die konfirmatorischen Faktorenanalysen – unabhängig von der Anzahl der modellierten Faktoren – zufriedenstellende Fit-Werte aus. Allerdings zeigen χ^2 -Differenztests, dass das 3-Faktoren-Modell dem 2- und dem 1-Faktoren-Modell überlegen ist, was die strukturelle Validität der Skalen untermauert.

Führungsverhalten: Tab. 4 enthält die Ergebnisse der χ^2 -Differenztests zur Faktorenstruktur des wahrgenommenen Führungsverhaltens. Der ansteigende χ^2 -Wert deutet an, dass die Modelle mit absteigender Faktorenanzahl die Datenstruktur schlechter widerspiegeln. Dieser Abfall im Modellfit (unter Berücksichtigung der gewonnenen Freiheitsgrade) ist statistisch nur für das 2- und 1-Faktoren-Modell signifikant. Das bedeutet, dass ein 3-Faktoren-Modell die empirische Kovarianzmatrix am besten reproduziert. Auf der nachfolgend relevanten Analyseebene von Abteilungen bzw. Fachbereichen sind allerdings sehr hohe Zusammenhänge zwischen diesen Faktoren zu beobachten ($r_{SEVB-IU}=0,907$, $p<0,000$; $r_{SEVB-PF}=0,938$, $p<0,000$; $r_{IU-PF}=0,946$, $p<0,000$). Daher greifen wir in den nächsten Analyseschritten (Abschn. 4.2) auf das 1-Faktoren-Modell (Globalfaktor-Modell) zurück. Dies ist aus zwei weiteren Gründen sinnvoll: Erstens führen die hohen Faktorkorrelationen in allen Modellierungsvarianten des Führungsverhaltens (insbes. im 4-Faktoren-Modell; siehe Tab. 8 im Anhang) in multiplen Regressionen zu Multikollinearität, die inhaltlich aussagekräftige Ergebnisse verhindert. Zweitens ist aufgrund der kleinen Clusteranzahl ein weniger komplexes Modell vorzuziehen. Daher wurden einfache Regressionen geschätzt, um Erklärungsbeiträge des Globalfaktors *Unterstützende Führung* für Kerndimensionen der PLG-Arbeit zu ermitteln.

4.2 Vorhersage der PLG-Kerndimensionen durch das Führungsverhalten schulischer Leitungskräfte

Tab. 5 zeigt, dass die Dimensionen *Normativer Konsens* und *Tragfähige Infrastruktur* überzufällig und mit mittlerer bis hoher Effektstärke durch den Globalfaktor *Unterstützende Führung* vorhergesagt werden können (NK: $\beta_g=0,481$, $p=0,014$; TI: $\beta_g=0,650$, $p=0,000$). Auch für die Dimension *Kooperative Entwicklung* kann ein moderater Zusammenhang beobachtet werden, der aber statistisch nicht signifikant ausfällt (KE: $\beta_g=0,314$, $p=0,206$). Angesichts der kleinen Clusteranzahl auf Fachbereichsebene ist dabei die geringe Teststärke der Stichprobe zur Identifikation von statistisch signifikanten Koeffizienten zu berücksichtigen. Die in Anhang A6 dargestellten Ergebnisse einer Simulationsstudie zeigen, dass die Wahrscheinlichkeit bei 54% liegt, aufgrund von $p=0,206$ zu schließen, dass der Effekt des Führungsverhaltens auf die Kooperative Entwicklung in der PLG-Arbeit ($\beta=0,314$) nicht vorliegt, obwohl dieser tatsächlich vorliegt und nachweisbar wäre, hätte man eine ausreichend große Stichprobe (gegeben einem Signifikanzniveau von $\alpha=0,05$).

Tab. 3 Prüfung der Faktorensstruktur PLG-Kerndimensionen mittels χ^2 -Differenztests für genesete Modelle

M	Par	χ^2	df	χ^2/df	AIC	BIC	CFI	TLI	RMSEA	SRMR _{L1}	SRMR _{L2}	$\Delta\chi^2$	Δdf	p
3	92	561,94	391	1,44	23.085,04	23.451,10	0,936	0,931	0,033	0,058	0,517	–	–	–
2	90	594,67	393	1,51	23.113,77	23.471,87	0,925	0,919	0,036	0,060	0,591	32,73	2	0,000
1	89	609,88	394	1,55	23.126,99	23.481,11	0,919	0,914	0,037	0,061	0,606	15,22	1	0,000

M geschätzte Modelle (die Zahl gibt die Anzahl der modellierten Faktoren an), Par Parameter, χ^2 Chi-Quadrat, AIC Akaikes Informationskriterium, BIC Bayessches Informationskriterium, CFI Comparative Fit Index, TLI Tucker Lewis Index, RMSEA Root Mean Square Error of Approximation, SRMR_{L1/L2} Standardized Root Mean Square Residual auf Level 1 (Lehrerebene) und Level 2 (Fachbereichsebene), $\Delta\chi^2$ Veränderung im χ^2 -Wert, Δdf Veränderung in den Freiheitsgraden, p Signifikanzniveau
 N für alle Modelle = 395

Tab. 4 Prüfung der Faktorenstruktur des wahrgenommenen Führungsverhaltens mittels χ^2 -Differenztests für genestete Modelle

M	Par	χ^2	df	χ^2/df	AIC	BIC	CFI	TLI	RMSEA	SRMR _{L1}	SRMR _{L2}	$\Delta\chi^2$	Δdf	p
4	64	246,09	160	1,54	13.544,39	13.798,38	0,968	0,964	0,037	0,045	0,379	–	–	–
3	61	247,71	163	1,52	13.540,00	13.782,09	0,969	0,965	0,036	0,046	0,383	1,62	3	0,655
2	59	255,30	165	1,55	13.543,59	13.777,74	0,967	0,963	0,037	0,047	0,396	7,59	2	0,022
1	58	283,29	166	1,71	13.569,58	13.799,76	0,957	0,953	0,043	0,049	0,423	27,99	1	0,000

M geschätzte Modelle (die Zahl gibt die Anzahl der modellierten Faktoren an), Par Parameter, χ^2 Chi-Quadrat, df Freiheitsgrade, AIC Akaikes Informationskriterium, BIC Bayessches Informationskriterium, CFI Comparative Fit Index, TLI Tucker Lewis Index, RMSEA Root Mean Square Error of Approximation, SRMR_{L1/L2} Standardized Root Mean Square Residual auf Level 1 (Lehrerebene) und Level 2 (Fachbereichsebene), $\Delta\chi^2$ Veränderung im χ^2 -Wert, Δdf Veränderung in den Freiheitsgraden, p Signifikanzniveau

N für alle Modelle = 395

Tab. 5 Doppelt-latente Zweiebenen-Strukturgleichungsmodelle zur Vorhersage der PLG-Dimensionen durch das Führungsverhalten

	Modell 1 AV: KE	Modell 2 AV: NK	Modell 3 AV: TI	Modell 4 AV: KE/NK/TI
UV: UF β_{UF}	0,314	0,481	0,650	0,254/0,430/0,651
p	0,206	0,014	0,000	0,330/0,032/0,000
R^2	0,10	0,23	0,42	0,07/0,19/0,42
Par	45	41	44	84
χ^2	268,23	231,14	263,99	523,67
df	98	79	99	276
χ^2/df	2,74	2,93	2,67	1,90
AIC	10.383,44	9168,21	9514,20	16598,99
BIC	10.562,49	9331,35	9689,27	16933,21
CFI	0,918	0,919	0,932	0,928
TLI	0,908	0,908	0,924	0,920
RMSEA	0,066	0,070	0,065	0,048
SRMR _{L1}	0,057	0,054	0,057	0,061
SRMR _{L2}	0,289	0,251	0,231	0,356

AV abhängige Variable, UV unabhängige Variable, KE Kooperative Entwicklung, NK Normativer Konsens, TI Tragfähige Infrastruktur, UF Unterstützende Führung, β_{UF} Regressionskoeffizient des Globalfaktors UF auf Fachbereichsebene, p Signifikanzniveau, R^2 Varianzaufklärung auf Fachbereichsebene, Par Parameter, χ^2 Chi-Quadrat, df Freiheitsgrade, AIC Akaiikes Informationskriterium, BIC Bayessches Informationskriterium, CFI Comparative Fit Index, TLI Tucker Lewis Index, RMSEA Root Mean Square Error of Approximation, SRMR_{L1/L2} Standardized Root Mean Square Residual auf Level 1 (Lehrerebene) und Level 2 (Fachbereichsebene)

N für alle Modelle = 395

Diese Fehlerwahrscheinlichkeit liegt deutlich über der tolerierbaren Obergrenze von 20 % (Cohen 1988). Die quadrierten Effektstärken (R^2) lassen hingegen erkennen, dass zwischen 10 und 42 % der Varianz der auf Fachbereichsebene geteilten Einschätzungen von PLG-Aktivitäten durch das auf Fachbereichsebene wahrgenommene Führungsverhalten der Schulleiter/innen erklärt werden kann. Werden die abhängigen Variablen KE, NK und TI simultan geschätzt, ändern sich diese Befunde kaum.

5 Schlussbetrachtung

5.1 Erkenntnisgewinne, Limitationen und weiterer Forschungsbedarf

Vor dem Hintergrund erweiterter einzelschulischer Gestaltungsräume und parallel wachsender Gestaltungsaufgaben der Schulleitung befasste sich die hier berichtete Studie mit der Frage, welchen Beitrag schulische Führungskräfte zur Förderung der Lehrendenkooperation in Fachbereichen nach dem Vorbild professioneller Lerngemeinschaften (PLGen) leisten können. Hierzu wurden Befragungsdaten von 395 Berufsschullehrkräften in 47 berufsfachlich organisierten Abteilungen mittels Mehrebenen-Strukturgleichungsmodellen ausgewertet. Den Analysen gingen konzeptionelle

Überlegungen und strukturprüfende Verfahren voran, die angesichts einer erheblichen Operationalisierungsvielfalt in der PLG-Forschung dazu diene, wesentliche Elemente sowohl der professionellen Zusammenarbeit als auch der sie begünstigenden Führungspraktiken einzugrenzen.

Die Literatursichtung ergab eine Unterscheidung von drei Kerndimensionen der PLG-Arbeit (*Kooperative Entwicklung*, *Normativer Konsens* und *Tragfähige Infrastruktur*) sowie vier Ansatzpunkten unterstützender Führung (*Strukturelle Erleichterungen*, *Impulse für die Unterrichtsentwicklung*, *Vertrauens- und Beziehungsstiftung* sowie *Partizipative Führung*). Anschließend spezifizierte *doppelt latente Strukturgleichungsmodelle* legen allerdings nahe, dass die auf Fachbereichsebene geteilten Einschätzungen dieser Aspekte das theoretisch angenommene Differenzierungsniveau nicht ganz erreichen. Zwar spiegelt ein theoriekonformes 3-Faktoren-Modell für Kerndimensionen der PLG-Arbeit die Antwortmuster der befragten Berufsschullehrkräfte am besten wider; bei den Führungspraktiken lassen sich hingegen nur drei statt vier postulierte Dimensionen empirisch abbilden. Zudem korrelieren die Führungspraktiken in den geteilten Wahrnehmungen der Fachbereichsmitglieder hoch miteinander, so dass zur Vermeidung von Multikollinearitätsproblemen in den sich anschließenden Regressionen ein Globalfaktor *Unterstützende Führung* herangezogen wurde. Interessanterweise findet sich die Verwendung eines Globalfaktors – größtenteils unkommentiert – auch in anderen Studien der bedingungsbezogenen PLG-Forschung (z. B. Bryk et al. 1999; Li et al. 2016; Webs und Holtappels 2018). Als Desiderat lassen sich deshalb Instrumentenentwicklungen und Validierungsstudien benennen, die nicht nur *generische* Führungsstrategien in schulischen Organisationen (z. B. Brauckmann und Pashiardis 2011), sondern auch spezifische Maßnahmen zur gezielten Förderung professioneller Lerngemeinschaften differenzierter abbilden als es in bisherigen Untersuchungen gelingt.

Die durchgeführten Regressionen mit dem Globalfaktor der *unterstützenden Führung* entsprechen dennoch unserer forschungsleitenden Annahme, dass hohe Ausprägungen der PLG-Kerndimensionen vorrangig in solchen Abteilungen zu finden sind, deren Arbeit auf vielfältige Weise begünstigt wird (siehe hierzu vertieft 5.2). Dies gilt vor allem für die *ideellen* und *infrastrukturellen* Dimensionen der PLG-Arbeit als Kriteriumsvariablen, deren Variationen statistisch überzufällig und mit mittleren bis starken Effektstärken durch ein unterstützendes Führungsverhalten aufgeklärt werden können. Stichprobenspezifische, jedoch praktisch bedeutsame Erklärungsbeiträge ergeben sich zudem für die *aktionale Dimension* der PLG-Arbeit. Dass die hier erhobenen Führungspraktiken die Intensität von PLG-Aktivitäten ursächlich hervorbringen, kann mit dem querschnittlichen Design unserer Studie freilich nicht nachgewiesen werden. Die methodisch fortgeschrittenen Analyseverfahren (doppelt-latente Mehrebenen-Strukturgleichungsmodellierung zur Kontrolle von Mess- und Stichprobenfehlern, vergleichende Zusatzanalysen nach Ausschluss von Clustern mit geringer Urteilerübereinstimmung, Simulationsstudien) legen aber nahe, dass eine intensive PLG-Arbeit in Fachbereichen nicht unabhängig vom Agieren der Schulleitung entsteht. Dennoch stellen längsschnittliche Untersuchungen ein weiteres gewichtiges Forschungsdesiderat dar, um kausale Wirkungsanalysen vornehmen zu können.

Mit Blick auf die Skalen zum Führungsverhalten ist hervorzuheben, dass die niedrige Reliabilität der Skalen auf Fachbereichsebene (Stichprobenfehler) die Analysen der Effekte des Führungsverhaltens trotz der verwendeten Doubly-Latent-Modelle erschwert. Aufgrund des Erhebungsdesigns war es nämlich nicht möglich die tatsächlichen Größen der Fachbereiche zu erfassen, sodass unklar ist, ob die statistische Kontrolle des Stichprobenfehlers in der vorliegenden Studie adäquat ist. Aber auch hier kann auf die Zusatzanalysen verwiesen werden, die zeigen, dass selbst nach Ausschluss von Clustern mit geringer Urteilerübereinstimmung die ermittelten Befunde konsistent bleiben.

Einschränkend ist weiterhin zu vermerken, dass Modellierungen mit einem Globalfaktor *Unterstützende Führung* zwangsläufig zulasten differenzierter Aussagen über die *relativen* Erklärungsbeiträge der hierin gebündelten Aktivitäten der Schulleitung gehen. Die mäßige empirische Bestätigung der aus umfangreichen Literatursichtungen gewonnenen Faktorenstruktur wurde bereits weiter oben moniert. Ein Ausweichen auf überprüfte Skalen zur Erfassung *allgemeiner* Führungsstrategien ohne Zuschnitt auf die Unterstützung von PLGen schien für die forschungsleitenden Fragestellungen des aktuellen Artikels allerdings ebenfalls suboptimal.

Eine letzte Limitation der vorliegenden Studie stellt der begrenzte Geltungsbereich der Ergebnisse dar. Einerseits verbietet die mangelnde Repräsentativität der Klumpenstichprobe eine Verallgemeinerung der Befunde. Andererseits verbietet der Untersuchungskontext beruflicher Schulen vorschnelle Generalisierungen auf andere Schularten, auch wenn unsere Literatursichtung eine hohe inhaltliche Konsistenz der postulierten unterstützenden Führungspraktiken, unabhängig von der betrachteten Schulart, nahelegen. Für die Übertragbarkeit spricht, dass sich bei der *Personalführung* – anders als im Bereich der Ziele und didaktischen Prinzipien des Unterrichtens – an beruflichen Schulen grundsätzlich ähnliche Fragen stellen wie an allgemeinbildenden Schulen. Relativierend sind wiederum unterschiedliche Qualifizierungshintergründe des Personals sowie unterschiedlich große Leitungsspannen zu berücksichtigen (die jedoch nicht nur schulart-, sondern auch schulgrößenabhängig variieren). Berücksichtigt man ferner, dass das berufliche Schulsystem in DE/AUT von 58%/78% aller Lernenden der Sekundarstufe II besucht wird (OECD 2010), stellt die Gewinnung wissenschaftlicher Erkenntnisse in diesem Sektor einen wichtigen Beitrag zur Qualitätssicherung des Bildungsangebotes dar. Dies gilt insbesondere für die Forschung zur professionellen Zusammenarbeit von Lehrkräften und den Möglichkeiten ihrer gezielten Förderung durch die Schulleitung, da hierzu noch kaum Befunde für den berufsbildenden Bereich vorliegen.

5.2 Implikationen für die schulische Führungspraxis

Erfolgsträchtige Stützstrategien für die Arbeit von PLGen sind *vielfältig und keineswegs einseitig* anzulegen (z. B. Borchers 2009; Cormier und Olivier 2009). Hierfür müssen nicht nur günstige *arbeitsorganisatorische* Voraussetzungen geschaffen werden, die in manchen Veröffentlichungen sogar priorisiert werden (z. B. Wahlstrom und Louis 2008); vielmehr dürfen auch *arbeitsklimatische* Bedingungen nicht vernachlässigt werden, wenn eine ernsthafte, produktive Auseinandersetzung über hochwertige Lehr-Lern-Prozesse in einer vertrauensvollen, angst- und vorwurfs-

freien Atmosphäre dauerhaft etabliert werden soll (z. B. Berkemeyer et al. 2015; Gray et al. 2016). Einem frühzeitigen Ende entsprechender Entwicklungsprozesse sollte die Schulleitung zudem (*mikro-)politisch* entgegenwirken, indem sie die domänenspezifische Expertise der PLG-Mitglieder anerkennt und in innerschulischen Entscheidungs- und Gestaltungsfragen berücksichtigt (z. B. Scribner et al. 2002). Um darüber hinaus auch die *inhaltliche* Ausrichtung professioneller Zusammenarbeit auf kontinuierliche unterrichtliche Verbesserungen zu gewährleisten, kommt schulischem Führungspersonal die wesentliche Aufgabe zu, sachdienliche Impulse und begründete Rückmeldungen zur Bestätigung oder aber Korrektur von Entwicklungsschritten zu geben. Vor allem diese Funktionen des Monitorings und Feedbacks werden jedoch im deutschsprachigen Raum von vielen Leitungskräften nur zögerlich umgesetzt, was Professionalisierungsbedarfe für das Leitungshandeln selbst nahelegt (Thiel et al. 2018).

Funding Open access funding provided by Johannes Kepler University Linz.

Anhang

A1 Stichprobe und Fragebogenitems

Stichprobenziehung Aus ökonomischen Gründen wurden *cluster samples* (Klumpenstichprobe, Bortz und Schuster 2010, S. 81) erfasst. Von einer vorab geplanten Stichprobenziehung wurde u. a. deshalb abgesehen, weil die Ziehung einer repräsentativen Stichprobe im vorliegenden Fall aus theoretischen Gründen nicht trivial ist: Aufgrund weitgehend fehlender Forschungsbefunde im deutschsprachigen Raum ist für die hier verfolgte Forschungsfrage bisher nicht geklärt, *welche* konkreten Merkmale von Lehrkräften oder Schulen als relevant für die Untersuchung von Schulleitungseffekten auf die Zusammenarbeit in PLGen gelten können. Dies wäre aber eine Voraussetzung für eine repräsentative Stichprobe (ebd., S. 80; Keller et al. 2017, S. 42). Aus u. a. diesen Gründen sind Klumpenstichproben in der Bildungsforschung nicht unüblich.

Generalisierbarkeit Die Generalisierbarkeit der Ergebnisse, die auf Basis von Klumpenstichproben gewonnen werden, hängt davon ab, wie stark sich die untersuchten Lehrpersonen von Schule zu Schule unterscheiden und wie gut die ausgewählten berufsbildenden Schulen die Population repräsentieren (ebd.). Wir haben daher die Verteilung der Variablen *Lehrende pro Schultyp*, *Lehrendengeschlecht* und *Lehrendenalter* in der vorliegenden Stichprobe mit den Verteilungen dieser Variablen in der Gesamtpopulation berufsbildender Lehrkräfte verglichen. Die Verteilungen der Gesamtpopulation ($N=14.013$ Lehrkräfte) sind öffentlich zugänglich³. Chi-Quadrat-Tests zeigen, dass die Klumpenstichprobe lediglich für das Lehrendengeschlecht repräsentativ ist. Allerdings sind die Abweichungen in der Anzahl der Lehrenden pro Schultyp und Lehrendenalter auf wenige Kategorien beschränkt:

³ <https://www.statistik.bayern.de/veroeffentlichungen/>.

Lehrkräfte in Berufsfachschulen und Lehrkräfte zwischen 40–45 Jahren sind in unserer Stichprobe überrepräsentiert; Lehrkräfte zwischen 60–65 Jahren sind unterrepräsentiert.

Rücklaufquote Die Fragebögen wurden an Schulen gesandt, die bereits an einer Vorläuferstudie zu beruflichen Anforderungen von Schulleitungen teilgenommen und sich bereit erklärt hatten, an künftigen Studien ebenfalls mitzuwirken. Im ersten Anlauf reagierten lediglich 5 der 34 Schulen nicht; diese konnten aber durch eine nochmalige Aufforderung für die Studie gewonnen werden, sodass alle angefragten Schulen im Sommer 2011 auch an der Befragung teilnahmen (Rücklaufquote auf Schulebene = 100%). Da es den Schulen freigestellt war, welcher Fachbereich sich an der Befragung beteiligt und auch schulseitig großer Wert auf eine möglichst weitreichende Anonymisierung gelegt wurde, wurde auf Zusatzinformationen wie die Größe der Fachbereiche verzichtet. Daher ist eine Rücklaufquote auf Lehrkräfteebene nicht bestimmbar. Die Thematik der Fachbereichsgrößen wird in den Bereichen A5 und A6 in diesem Anhang nochmals aufgegriffen.

Führungsdimensionen

SE = Strukturelle Erleichterungen

Die Schulleitung...

- ...unterstützt die Lehrkräfte durch eine gelungene Unterrichtsorganisation. (P3)
- ...versucht, die Lehrkräfte von allen nicht unmittelbar unterrichtsbezogenen Aufgaben zu entlasten. (P4)
- ...schafft effiziente Arbeitsstrukturen für die Bewältigung der täglichen Aufgaben der Kollegiumsmitglieder. (P4)

IU = Impulse für Unterrichtsentwicklung

Die Schulleitung...

- ...diskutiert regelmäßig im Fachkollegium Fragen der Unterrichtsgestaltung. (P6)
- ...führt regelmäßige Unterrichtsbesuche durch, um sich ein Bild von der Qualität der pädagogischen Arbeit zu machen. (P7)
- ...erarbeitet mit den Lehrkräften eine gemeinsame Vorstellung von gutem Unterricht. (P7)

VB = Vertrauens- und Beziehungsstiftung

Die Schulleitung...

- ...schafft entspannte Arbeitssituationen, die auf gegenseitigem Vertrauen basieren. (P1)
- ...lebt Handlungsprinzipien vor, an denen sich unsere gemeinsame Arbeit ausrichtet. (P1)
- ...sorgt bei Auseinandersetzungen für Vermittlung und sozialen Ausgleich. (P2)

- ...schafft eine von gemeinsamen Idealen getragene Schulgemeinschaft. (P2)
- ...richtet ihr Handeln an übergeordneten Werten und Normen aus. (P3)

PF = Partizipative Führung

Die Schulleitung...

- ...bindet die Lehrkräfte in verantwortungsvolle Führungsaufgaben ein. (P5)
- ...beteiligt das Fachkollegium an der Schulgestaltung. (P5)
- ...berät alle wichtigen Entscheidungen vorab mit Experten aus dem Kollegium. (P6)

PLG-Dimensionen

KE =Kooperative Entwicklung

- Selbst- und gegenseitige Evaluationen sind selbstverständliche Teile unserer Arbeit. (P10)
- Bei uns finden gegenseitige Unterrichtsbesuche regelmäßig statt.
- In unserem Fachbereich sind gemeinsame Planungen von Unterrichtsthemen eine Ausnahme.* (P8)
- Die Lehrkräfte führen bei uns häufig gemeinsame Projekte durch. (P8)
- Wir bereiten Unterrichtseinheiten häufig gemeinsam vor. (P9)
- In unserem Fachbereich kommt es selten vor, dass Unterrichtsthemen gemeinsam aus mehreren Perspektiven entwickelt werden.* (P9)
- Bei uns hat man in der Regel keine Ahnung davon, was andere Kollegen gerade im Unterricht behandeln.* (P10)

NK =Normativer Konsens

- Jeder Kollege hat seine eigene Vorstellung davon, wie Unterricht zu gestalten wäre.* (P11)
- Die didaktischen Vorstellungen im Fachkollegium lassen sich kaum auf einen gemeinsamen Nenner bringen.* (P119)
- Bei einem Wechsel des Fachlehrers müssen sich die Schüler auf andere Lehrstrategien einstellen.* (P12)
- Bei uns können die Schüler davon ausgehen, dass alle Lehrkräfte ähnliche Anforderungen im Unterricht stellen. (P12)
- Bei uns im Fachkollegium arbeitet jeder so, wie er es selber für richtig hält.*

TI = Tragfähige Infrastruktur

- Bei uns hat man oft das Gefühl, niemand fühlt sich für Aufgaben verantwortlich.* (P13)
- In unserem gemeinsamen Verantwortungsbereich trifft die Belastung durch schulische Aufgaben immer nur einige wenige Lehrkräfte.* (P13)
- Bei uns im Fachbereich ist klar, wer wofür zuständig ist. (P14)
- Bei uns sind Projektarbeiten gut organisiert. (P15)

- Man kann man in unserem Fachbereich sicher sein, dass Entscheidungen auch konsequent umgesetzt werden. (P15)
- Unsere Arbeit besteht aus vielen zusammenhanglosen Einzelaktionen.* (P14)
- Alle Fachkollegen sind ausreichend über wichtige Entscheidungen und Neuerungen informiert. (P16)
- Bei uns sind Lehrerkonferenzen gut geplant und durchgeführt. (P16)
- Unsere Arbeitsabläufe sind gut strukturiert. (P16)

Anmerkungen: * = invers; P1–P16 = Parcels in den PLG-Kerndimensionen und im Globalfaktor Unterstützende Führung.

A2 Latente Korrelationen

Tab. 6 Latente Korrelationen zwischen den untersuchten Führungs- und PLG-Dimensionen auf Lehrerebene

	KE	NK	TI	SE	IU	VB	PF
KE	1	–	–	–	–	–	–
NK	0.620	1	–	–	–	–	–
TI	0.401	0.677	1	–	–	–	–
SE	0.282	0.429	0.735	1	–	–	–
IU	0.531	0.493	0.558	0.660	1	–	–
VB	0.355	0.551	0.788	0.790	0.706	1	–
PF	0.415	0.486	0.704	0.744	0.753	0.851	1

KE Kooperative Entwicklung, *NK* Normativer Konsens, *TI* Tragfähige Infrastruktur, *SE* Strukturelle Erleichterungen, *IU* Impulse für Unterrichtsentwicklung, *VB* Vertrauens- und Beziehungsstiftung, *PF* Partizipative Führung

Alle latenten Korrelationen sind hoch signifikant ($p < 0,001$)

Tab. 7 Latente Korrelationen zwischen den untersuchten PLG-Dimensionen auf Fachbereichsebene

	KE	NK	TI
KE	1	–	–
NK	0.848	1	–
TI	0.903	0.795	1

KE Kooperative Entwicklung, *NK* Normativer Konsens, *TI* Tragfähige Infrastruktur

Alle latenten Korrelationen sind hoch signifikant ($p < 0,001$)

Tab. 8 Latente Korrelationen zwischen den untersuchten Führungsdimensionen auf Fachbereichsebene

	SE	IU	VB	PF
SE	1	–	–	–
IU	0.880	1	–	–
VB	0.992	0.920	1	–
PF	0.922	0.946	0.943	1

SE Strukturelle Erleichterungen, *IU* Impulse für Unterrichtsentwicklung, *VB* Vertrauens- und Beziehungsstiftung, *PF* Partizipative Führung

Alle latenten Korrelationen sind hoch signifikant ($p < 0,001$)

A3 Technische Hinweise

Bildung von Itemparcels

Bei kleinen Stichproben wie der vorliegenden (Clusteranzahl=47), empfehlen Little et al. (2002) im Rahmen der Messmodelle Itemparcels zu bilden. Dies hat den Vorteil, dass weniger Parameter geschätzt werden müssen, d.h. die Modellkomplexität wird reduziert. Sparsamere Messmodelle wiederum können zu einem höheren Modellfit führen, da u. a. weniger Korrelationen zwischen den Itemresiduen möglich sind (Little et al. 2002, S. 155). Insgesamt kann damit die Datenstruktur besser wiedergespiegelt werden. Aus diesen Gründen werden jeweils zwei Items zu einem Parcel zusammengefasst. In Anhang A1 ist ersichtlich, welche Itempaare zu einem Itemparcel zusammengefasst wurden.

Konfigurale Messinvarianz

Um die lehrerindividuellen Wahrnehmungen der PLG-Arbeit auf Fachbereichsebene als geteilte Wahrnehmung der PLG-Arbeit interpretieren zu können, sind sogenannte cross-level constraints nötig (Stapleton et al. 2016; Jak und Jorgensen 2017). Mit cross-level constraints ist hier die Gleichsetzung der Faktorladung eines Items auf Individual- und Gruppenebene gemeint. Damit sind auf beiden Ebenen nicht nur die Faktorenstruktur, sondern auch die Faktorenladungen je Item (bzw. Itemparcel) identisch. Die so erhaltene konfigurale Messinvarianz (über die Analyseebenen hinweg) erlaubt es, die Faktoren verschiedener Analyseebenen als die Individual- und Gruppenkomponente derselben Variable zu interpretieren (Jak und Jorgenson 2017, S. 3). Über die konfigurale Messinvarianz hinaus (Fixierung der Faktorladungen über die Levels hinweg) ist dagegen eine Berücksichtigung der skalaren Messinvarianz (Fixierung der Itemintercepts über die Levels hinweg) nicht möglich: „Equality of intercepts, on the other hand, cannot be tested across levels because the intercepts apply only to the observed variables, not separately for within- and between-level components. The common practice of fixing factor means to zero for identification of the mean structure makes it easy to show that within-level intercepts are expected to be zero“ (Jak und Jorgenson 2017, S. 8). Damit wird auch der Unterschied zwischen (a) der Überprüfung der Messinvarianz zwischen Gruppen und (b) der Überprüfung der Messinvarianz zwischen Ebenen (Levels) deutlich.

A4 Prüfung der Dimensionsstruktur

Da es sich bei den eingesetzten Skalen zur Arbeit in PLGen und zum Führungsverhalten von Schulleitungen nicht um standardisierte Messinstrumente handelt, wurde die Faktorenstruktur durch eine schrittweise Reduktion der Anzahl der hinter den Items vermuteten latenten Konstrukte geprüft.

Für das Führungsverhalten wurden vier genestete Modelle spezifiziert (siehe Abb. 1).

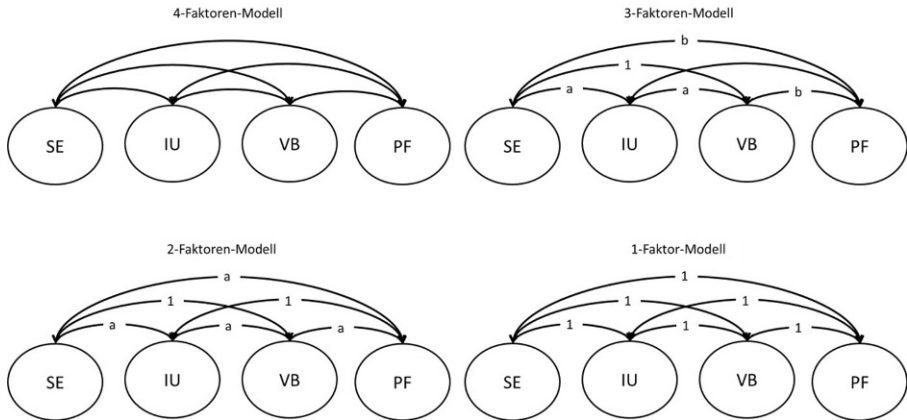


Abb. 1 Führungsverhalten: Spezifikation der Faktorenmodelle auf Fachbereichsebene (Level 2). *SE* Strukturelle Erleichterungen, *IU* Impulse für Unterrichtsentwicklung, *VB* Vertrauens- und Beziehungsstiftung, *PF* Partizipative Führung. Auf Lehrendenebene (Level 1) wurden dieselben Modelle allerdings ohne die abgebildeten Restriktionen spezifiziert

4-Faktoren-Modell: Alle Korrelationen zwischen den latenten Konstrukten auf Fachbereichsebene wurden frei geschätzt.

3-Faktoren-Modell: Die latente Korrelation zwischen den am stärksten zusammenhängenden Teilkonstrukten *SE* und *VB* wurde auf den Wert 1 fixiert (perfekte Korrelation⁴). Darüber hinaus wurden die Korrelationen von *SE* und *VB* mit *IU* (*a*-Pfade) und *PF* (*b*-Pfade) gleichgesetzt. Dadurch ergeben sich 3 Faktoren.

2-Faktoren-Modell: Zusätzlich zur *SE*-*VB*-Korrelation (siehe 3-Faktoren-Modell) wurde die latente Korrelation zwischen den beiden Konstrukten *IU* und *PF* auf 1 fixiert. Alle anderen latenten Korrelationen wurden gleichgesetzt (*a*-Pfade). Dadurch ergeben sich 2 Faktoren.

1-Faktor-Modell: Alle latenten Korrelationen wurden auf 1 fixiert.

Analog wurde die Dimensionsstruktur der *PLG*-Skalen geprüft (siehe Abb. 2).

3-Faktoren-Modell: Alle Korrelationen zwischen den latenten Konstrukten auf Fachbereichsebene wurden frei geschätzt.

2-Faktoren-Modell: Die latente Korrelation zwischen den am stärksten zusammenhängenden Teilkonstrukten *NK* und *TI* wurde auf den Wert 1 fixiert (perfekte Korrelation). Darüber hinaus wurden die Korrelationen von *TI* und *NK* mit *KE* gleichgesetzt (*a*-Pfade). Dadurch ergeben sich 2 Faktoren.

1-Faktor-Modell: Alle latenten Korrelationen wurden auf 1 fixiert.

⁴ Bei allen Modellen wurden die Varianzen der latenten Konstrukte auf 1 fixiert, um so die Metrik der Skala zu identifizieren.

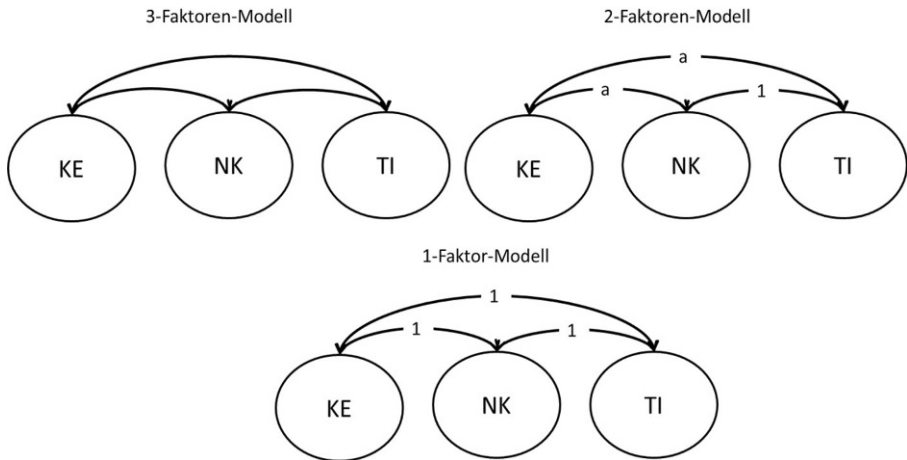


Abb. 2 PLG-Dimensionen: Spezifikation der Faktorenmodelle auf Fachbereichsebene (Level 2). *KE* Kooperative Entwicklung, *NK* Normativer Konsens, *TI* Tragfähige Infrastruktur. Auf Lehrendenebene (Level 1) wurden dieselben Modelle allerdings ohne die abgebildeten Restriktionen spezifiziert

A5 Zusatzanalysen nach Ausschluss von Fachbereichen mit geringer Urteilerübereinstimmung

Da der Sampling Error (ICC(2)) – wenn auch in den Analysen mitmodelliert – die Validität der Prädiktorvariable (Globalfaktor Unterstützende Führung, UF) beeinträchtigen könnte, haben wir entsprechend der von Lüdtke et al. (2006) empfohlenen Vorgehensweise Zusatzanalysen durchgeführt, die zeigen, inwiefern sich die untersuchten Effekte ändern, wenn man jene Cluster (= Fachbereiche) entfernt, die eine geringe Urteilerübereinstimmung (deutlich variierende Lehrerangaben zum Schulleiterverhalten) aufweisen. Diese Zusatzanalysen gliedern sich in 4 Schritte.

Erstens, für jeden Fachbereich wurde der $AD_{M(j)}$ -Index und $R_{WG(j)}$ -Index (nach Lüdtke et al. 2006) für die 4 Skalen zum wahrgenommenen unterstützenden Führungsverhalten (UF) berechnet. Es zeigte sich, dass 18 der 47 Fachbereiche in mindestens einer Skala und mindestens einem Index das Cut Off-Kriterium (<1 für den $AD_{M(j)}$ -Index⁵, $>0,7$ für den $R_{WG(j)}$ -Index) nicht erreichen. Diese Fachbereiche wurden in allen Zusatzanalysen ausgeschlossen, sodass eine Teilstichprobe mit 29 Fachbereichen und 248 Lehrer/innen gewonnen wurde.

Zweitens, für die Teilstichprobe wurde die Cluster-Level-Reliabilität ICC(2) für den Globalfaktor UF berechnet. Der ICC(2)-Wert ist nach Ausschluss der 18 Fachbereiche von 0,48 auf 0,63 gestiegen – auch der ICC(1) des Globalfaktors ist von 0,10 auf 0,17 gestiegen. Ob der Sampling Error bei einem ICC(2) von 0,63 ausreichend gering ist, um von einer akzeptablen Validität unseres Prädiktors ausgehen zu können, kann abschließend nicht beurteilt werden, da es sich beim Cut Off von 0,70 lediglich um eine Daumenregel handelt (Lüdtke et al. 2006). Letztlich muss die Va-

⁵ Nach Lüdtke et al. (2006) liegt der Cut-Off bei $A/6$, wobei A die Anzahl der Antwortkategorien darstellt. Da die Antwortskala 6 Kategorien aufweist, ergibt sich der Wert 1.

Tab. 9 Zweiebenen-Strukturgleichungsmodell zur Vorhersage der PLG-Dimensionen durch das Führungsverhalten in einer Teilstichprobe (nach Ausschluss von Fachbereichen mit geringer Urteilerübereinstimmung)

	Modell 1 AV: KE	Modell 2 AV: NK	Modell 3 AV: TI	Modell 4 AV: KE/NK/TI
UV: UF β_{UF}	0,319	0,412	0,476	0,249/0,408/0,463
p	0,385	0,082	0,02	0,472/0,084/0,025
R^2	0,10	0,17	0,23	0,06/0,17/0,21
Par	45	40	44	80
χ^2	169,71	148,28	18,95	397,92
df	98	80	99	279
AIC	6140	5364	5616	9947
BIC	6298	5505	5771	10232
CFI	0,933	0,927	0,931	0,935
TLI	0,925	0,918	0,923	0,928
RMSEA	0,057	0,058	0,058	0,041
SRMR _{L1}	0,064	0,063	0,064	0,066
SRMR _{L2}	0,247	0,205	0,217	0,295

AV abhängige Variable, UV unabhängige Variable, KE Kooperative Entwicklung, NK Normativer Konsens, TI Tragfähige Infrastruktur, UF Unterstützendes Führungsverhalten, β_{UF} Regressionskoeffizient des Globalfaktors auf Fachbereichsebene, p Signifikanzniveau, R^2 Varianzaufklärung auf Fachbereichsebene, Par Parameter, χ^2 Chi-Quadrat, df Freiheitsgrade, AIC Akaiikes Informationskriterium, BIC Bayessches Informationskriterium, CFI Comparative Fit Index, TLI Tucker Lewis Index, RMSEA Root Mean Square Error of Approximation, SRMR_{L1/L2} Standardized Root Mean Square Residual auf Level 1 (Lehrerebene) und Level 2 (Fachbereichsebene)

N für alle Modelle = 248

liditätsfrage auch vor dem Hintergrund der Itemformulierungen diskutiert werden. Hier zeigt die umfangreiche theorie- und literaturbasierte Darstellung im Manuskript, dass die Skalen relevante Aspekte der Schulleiterunterstützung inhaltlich abbilden.

Drittens wurden die Effekte des Globalfaktors UF auf die PLG-Dimensionen, wie sie in Tab. 5 im Manuskript berichtet sind, für die Teilstichprobe (d.h. ohne die 18 Fachbereiche mit geringer Urteilerübereinstimmung) erneut berechnet und in Tab. 9 dargestellt. Die Koeffizienten auf Basis der Teilstichprobe (siehe Tab. 9) sind zwar leicht niedriger als jene auf Basis der 47-Fachbereiche-Stichprobe (Tab. 5 im Manuskript), dennoch stützen sie die im Manuskript getätigten Aussagen bzw. Interpretationen. Die Interpretation, dass das wahrgenommene Schulleiterverhalten moderat positiv die NK- und TI-Dimension der Lehrerverkooperationsarbeit vorher-sagt und schwach positiv die KE-Dimension vorhersagt, ändert sich also auch nach Ausschluss der Fachbereiche mit geringen Urteilerübereinstimmungen nicht.

Viertens, da die Stichprobe auf Fachbereichsebene nun noch kleiner geworden ist (29 statt 47 nach Clusterausschluss), wurden Simulationsstudien durchgeführt, um der Frage nachzugehen, ob die in Tab. 9 berichteten Effekte und Signifikanzwerte verzerrt sind. Die Simulation wurde wie folgt spezifiziert:

- Das Analysemodell und das Datengenerationsmodell sind identisch.
- Für die Spezifikation des Modells zur Datengeneration wurden die Parameter herangezogen, die sich auf Basis der Daten und des Analysemodells ergeben. Für das Analysemodell wurden dieselben Parameter als wahre Populationswerte verwendet.
- Die Simulation wurde mit 1000 Replikationen durchgeführt.
- Die Anzahl der Cluster wurde auf 29 mit je 9 Lehrpersonen ($N=261$) gesetzt.
- Als Evaluationskriterien wurden der relative Bias und die Power herangezogen. Der relative Bias beschreibt die Abweichung des über alle simulierten Replikationen gemittelten Parameters vom wahren Populationswert. Der Faustregel von Muthén und Muthén (2002; Hoogland und Boomsma 1998) zufolge sollte der relative Bias 10% des wahren Populationswertes nicht übersteigen. Die Power gibt die Wahrscheinlichkeit an, einen Typ-II-Fehler nicht zu begehen. Der Typ-II-Fehler beschreibt die Wahrscheinlichkeit, aufgrund von nicht-signifikanten Koeffizienten zu schließen, dass ein Effekt nicht vorliegt, obwohl dieser vorliegt und als signifikant identifizierbar wäre, wenn u. a. eine ausreichend große Stichprobe verfügbar wäre. Einer Daumenregel nach sollte die Power bei 0,80 liegen (Cohen 1988).

Tab. 10 enthält die Ergebnisse der Simulationsstudie. Die Ergebnisse zeigen, dass der relative Bias im akzeptablen Bereich liegt ($<10\%$; Muthén und Muthén 2002; Hoogland und Boomsma 1998). Dies deckt sich mit den Befunden der Simulationsstudie von Helm (2018). Die Ergebnisse von Helm (2018) zeigen, dass bei einer Clusteranzahl von 30 und einem ICC(1) von 0,15/0,20 der relative Bias bei 15%/5% liegt. Der UF-ICC(1) von 0,17 liegt in der Mitte, d. h. auch ohne die hier berichteten Simulationsstudien wäre anzunehmen, dass der relative Bias noch vertretbar ist (d. h., unter oder um die 10% liegt). Dies deckt sich mit den Ergebnissen in Tab. 10. Allerdings zeigt Tab. 10 auch, dass für alle Effekte des Führungsverhaltens (UF) auf

Tab. 10 Ergebnisse der Simulationsstudie (Teilstichprobe) für den Effekt des wahrgenommenen Führungsverhaltens auf die PLG-Dimensionen auf Fachbereichsebene

	Population	Average	M.S.E.	95% Cover	% Sig Coeff	Relativ Bias in %	RMSE
<i>Bivariate Modellierung</i>							
UF > KE	0,168	0,176	0,056	0,942	0,137	4,46	1,412
UF > NK	0,452	0,485	0,076	0,950	0,445	7,30	0,608
UF > TI	0,541	0,573	0,074	0,951	0,602	5,90	0,503
<i>Multivariate Modellierung</i>							
UF > KE	0,158	0,1701	0,0542	0,951	0,108	7,66	1,473
UF > NK	0,447	0,4854	0,0745	0,930	0,504	8,59	0,611
UF > TI	0,522	0,5578	0,0736	0,931	0,620	6,86	0,520

Es wurde die Teilstichprobe nach Ausschluss der Fachbereiche mit niedriger Urteilerübereinstimmung simuliert: $N=261$, Cluster $N=29$, durchschnittliche Cluster-Größe = 9

UF Führungsverhalten (Globalfaktor), KE Kooperative Entwicklung, NK Normativer Konsens, TI Tragfähige Infrastruktur

M.S.E. = Mean Square Error, Cover = Coverage, % Sig Coeff = Power = 1-Type-II-Fehler, RMSE = root mean square error

die PLG-Dimensionen (KE, NK, TI) die Power (Spalte „% Sig Coeff“) zu niedrig ist und daher die Signifikanzwerte in Tab. 9 nicht interpretiert werden sollten. Dennoch stützen die unverzerrten β -Koeffizienten die Befunde der Vollstichprobe (Tab. 5 im Manuskript).

Zusammenfassend ist festzuhalten, dass die hier berichteten Zusatzanalysen zeigen, dass die im Manuskript dargestellten Befunde bzw. die daraus abgeleiteten Schlüsse sich nicht wesentlich ändern, wenn die Analysen nur auf Basis von Fachbereichen durchgeführt werden, die über eine ausreichende Urteilerübereinstimmung (Lüdtke et al. 2006) verfügen. Damit ist auch die prädiktive Validität der Variable zum wahrgenommenen Führungsverhalten mit geringem ICC(2) untermauert. Das heißt, es gibt starke Belege dafür, dass sowohl die Datenlage, die Messinstrumente und das Design bzw. die Analyseverfahren geeignet sind, um die Forschungsfrage nach dem Zusammenhang zwischen wahrgenommenen Führungsverhalten und Lehrkooperationsmerkmalen zu beantworten.

A6 Simulationsstudie für die volle Stichprobe

Um der Frage nachzugehen, wie stark die untersuchten Effekte auf Fachbereichsebene sowie deren Signifikanzen durch den niedrigen ICC(1) und die geringe Anzahl an Fachbereichen tatsächlich verzerrt sind, haben wir für die im Manuskript in Tab. 5 berichteten Effekte des wahrgenommenen Führungsverhaltens auf die berichtete Arbeit in PLGen (KE, NK, TI) Monte Carlo Simulationsstudien durchgeführt. Diese Simulationsstudien wurden wie folgt spezifiziert:

- Das Analysemodell und das Datengenerationsmodell sind identisch.
- Für die Spezifikation des Modells zur Datengeneration wurden die Parameter der Modelle aus dem ursprünglichen Manuskript verwendet. Auch für das Analysemodell wurden die Parameter aus dem ursprünglichen Manuskript als wahre Populationswerte verwendet.
- Die Simulation wurde mit 1000 Replikationen durchgeführt.
- Die Anzahl der Cluster wurde auf 47 mit je 8 Lehrpersonen ($N=376$) gesetzt.
- Als Evaluationskriterien wurden der relative Bias und die Power herangezogen. Der relative Bias beschreibt die Abweichung des über alle simulierten Replikationen gemittelten Parameters vom wahren Populationswert. Der Faustregel von Muthén und Muthén (2002; Hoogland und Boomsma 1998) zufolge sollte der relative Bias 10% des wahren Populationswertes nicht übersteigen. Die Power gibt die Wahrscheinlichkeit an, einen Typ-II-Fehler nicht zu begehen. Der Typ-II-Fehler beschreibt die Wahrscheinlichkeit, aufgrund von nicht-signifikanten Koeffizienten zu schließen, dass ein Effekt nicht vorliegt, obwohl dieser vorliegt und als signifikant identifizierbar wäre, wenn u. a. eine ausreichend große Stichprobe verfügbar wäre. Einer Daumenregel nach sollte die Power bei 0,80 liegen.

Die Ergebnisse der Simulationsstudie sind in der nachstehenden Tab. 11 dargestellt. Die Ergebnisse zeigen, dass der sogenannte *relative Bias* im akzeptablen Bereich, d. h. unter 10% des wahren Populationswertes liegt (Muthén und Muthén 2002; Hoogland und Boomsma 1998). Diese Befunde decken sich mit den bei Zitzmann et al. (2016, S. 667; siehe auch Zitzmann et al. 2015, S. 695) berichteten

Tab. 11 Ergebnisse der Simulationsstudie (volle Stichprobe) für den Effekt des wahrgenommenen Führungsverhaltens auf die PLG-Dimensionen auf Fachbereichsebene

	Population	Average	M.S.E.	95% Cover	% Sig Coeff	Relativ Bias in %	RMSE
<i>Bivariate Modellierung</i>							
UF > KE	0,330	0,349	0,039	0,949	0,457	5,64	0,595
UF > NK	0,549	0,577	0,046	0,947	0,825	5,08	0,389
UF > TI	0,855	0,895	0,061	0,940	0,996	4,71	0,289
<i>Multivariate Modellierung</i>							
UF > KE	0,263	0,290	0,038	0,942	0,356	10,11	0,736
UF > NK	0,477	0,505	0,042	0,945	0,742	5,85	0,428
UF > TI	0,858	0,896	0,059	0,939	0,989	4,37	0,283

Es wurde die volle Stichprobe simuliert: $N = 376$, Cluster $N = 47$, durchschnittliche Cluster-Größe = 8
UF Führungsverhalten (Globalfaktor), *KE* Kooperative Entwicklung, *NK* Normativer Konsens, *TI* Tragfähige Infrastruktur

M.S.E. = Mean Square Error, *Cover* = Coverage, *% Sig Coeff* = Power = 1-Type-II-Fehler, *RMSE* = root mean square error

Befunden, die zeigen, dass in Stichproben mit einer niedrigen Zahl von Clustern ($N = 25$), kleinen Gruppengrößen ($N = 10$) und niedrigen ICC (= 0,10) der relative Bias bei ML-Schätzungen ebenfalls im akzeptablen Bereich (kleiner gleich 10%) liegt. Neben dem relativen Bias wird die Power (*% Sig Coeff*) in Tab. 11 angeführt. Auch dieses Evaluationskriterium liegt im Soll-Bereich, mit einer Ausnahme: Die Stichprobe und der Effekt des Führungsverhaltens (UF) auf die Kooperative Entwicklung (KE) sind zu klein, um für den angesprochenen Effekt eine ausreichende Power zu erreichen. Mit anderen Worten: Für diesen Effekt ($\beta = 0,330$) liegt die Wahrscheinlichkeit einen Typ-II-Fehler zu begehen, bei rund 0,54 ($1 - 0,457$), sie sollte aber nicht größer als 0,20 ($1 - 0,80$) sein. Das heißt, mit den vorliegenden Daten ist die Wahrscheinlichkeit relativ hoch, aufgrund des nicht-signifikanten Koeffizienten zu schließen, dass ein Effekt des Führungsverhaltens auf die Kooperative Entwicklung in der PLG-Arbeit nicht vorliegt, obwohl dieser tatsächlich vorliegt, hätte man eine ausreichend große Stichprobe. Bei allen anderen Effekten ist neben dem Koeffizienten auch die statistische Signifikanz interpretierbar.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz (<http://creativecommons.org/licenses/by/4.0/deed.de>) veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Literatur

- Achinstein, B. (2002). *Community, diversity, and conflict among schoolteachers. The ties that blind*. New York, London: Teacher College Press.
- Altrichter, H., & Helm, C. (2011). Schulentwicklung und Systemreform. In H. Altrichter & C. Helm (Hrsg.), *Akteure & Instrumente der Schulentwicklung* (S. 13–35). Baltmannsweiler: Schneider Verlag Hohengehren.
- Barton, R., & Stepanek, J. (2012). The impact of professional learning communities. *Principal's Research Review*, 7(4), 1–7.
- Berkemeyer, J., Berkemeyer, N., & Schwikal, A. (2015). Lernen als Leitbild. Internationale Erfahrungen zum Schulleitungshandeln im Kontext von Professionalisierungsprozessen von Lehrkräften. In J. Ber-

- kemeyer, N. Berkemeyer & F. Meetz (Hrsg.), *Professionalisierung und Schulleitungshandeln. Wege und Strategien der Personalentwicklung an Schulen* (S. 12–32). Weinheim, Basel: Beltz Juventa.
- Berkemeyer, N., Järvinen, H., Otto, J., & Bos, W. (2011). Kooperation und Reflexion als Strategien der Professionalisierung in schulischen Netzwerken. *Zeitschrift für Pädagogik*, 57, 225–247.
- Bolam, R., McMahon, A., Stoll, L., Thomas, S., & Wallace, M. (2005). *Creating and sustaining effective professional learning communities*. Research report, Bd. 637. London: DFES, University of Bristol.
- Bonsen, M., & Hübner, C. (2012). Unterrichtsentwicklung in professionellen Lerngemeinschaften. In K.-O. Bauer & N. Logemann (Hrsg.), *Effektive Bildung – Zur Wirksamkeit und Effizienz pädagogischer Prozesse* (S. 55–76). Münster: Waxmann.
- Bonsen, M., & Rolff, H. G. (2006). Professionelle Lerngemeinschaften von Lehrerinnen und Lehrern. *Zeitschrift für Pädagogik*, 52(2), 167–184.
- Borchers, B. T. (2009). *A Study to determine the practices of high school principals and central office administrators who effectively foster continuous professional learning in high schools*. Nicht veröffentlichte Dissertation. University of Minnesota
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Wiesbaden: Springer.
- Brauckmann, S. (2012). Schulleitungshandeln zwischen deconcentration, devolution und delegation (3D) – empirische Annäherungen aus internationaler Perspektive. *Empirische Pädagogik*, 26(1), 78–102.
- Brauckmann, S., & Pashiardis, P. (2011). A validation study of the leadership styles of a holistic leadership theoretical framework. *International Journal of Educational Management*, 25(1), 11–32.
- Bryk, A., Camburn, E., & Louis, K. S. (1999). Professional community in Chicago elementary schools: facilitating factors and organizational consequences. *Educational Administration Quarterly*, 35(5), 751–781.
- Buhren, C. G. (2015). Personalentwicklung durch kollegiale Hospitation und Professionelle Lerngemeinschaften. In N. Berkemeyer, J. Berkemeyer & F. Meetz (Hrsg.), *Professionalisierung und Schulleitungshandeln. Wege und Strategien der Personalentwicklung an Schulen* (S. 160–184). Weinheim: Beltz Juventa.
- Carpenter, D. (2015). School culture and leadership of professional learning communities. *International Journal of Educational Management*, 29(5), 682–694.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.
- Cormier, R., & Olivier, D. F. (2009). *Professional learning committees: characteristics, principals, and teachers*. Annual Meeting of the Louisiana Education Research Association, Lafayette.
- Cranston, J. (2011). Relational trust: the glue that binds a professional learning community. *Alberta Journal of Educational Research*, 57(1), 59–72.
- Ditton, H. Materialien – QualitätsSicherung in Schule und Unterricht. <http://www.quassu.net/seite4.htm>. Zugriffen: 3. Mai 2018.
- Dubs, R. (2009). Leitungsstrukturen in Bildungsorganisationen – Leadership und die Folgen für die Professionalität von Lehrenden. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus & R. Mulder (Hrsg.), *Lehrprofessionalität: Bedingungen, Genese, Wirkungen und ihre Messung* (S. 503–516). Weinheim: Beltz.
- Eaker, R., Dufour, R., & Dufour, R. (2002). *Getting started: reculturing schools to become professional learning communities*. Bloomington: Solution Tree.
- Fleming, G. L. (2004). Principals and teachers as continuous learners. In S. M. Hord (Hrsg.), *Learning together, leading together: changing schools through professional learning communities* (S. 20–30). New York: Teachers College Press.
- Fussangel, K., & Gräsel, C. (2014). Forschung zur Kooperation im Lehrerberuf. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf* (S. 846–864). Münster: Waxmann.
- Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychology Methods*, 19(1), 72–91.
- Gray, J., Kruse, S., & Tarter, C. J. (2016). Enabling school structures, collegial trust and academic emphasis: antecedents of professional learning communities. *Educational Management Administration & Leadership*, 44(6), 875–891.
- Hallam, P. R., Smith, H. R., Hite, J. M., Hite, S. J., & Wilcox, B. R. (2015). Trust and collaboration in PLC teams: teacher relationships, principal support, and collaborative benefits. *NASSP Bulletin*, 99(3), 193–216.
- Helm, C. (2018). How many classes are needed to assess effects of instructional quality? A Monte Carlo simulation of the performance of frequentist and Bayesian multilevel latent contextual models. *Psychological Test and Assessment Modeling*, 60(2), 265–285.

- Hipp, K. K., Huffman, J. B., Pankake, A. M., & Olivier, D. F. (2008). Sustaining professional learning communities: case studies. *Journal of Educational Change*, 9(2), 173–195.
- Hoogland, J. J., & Boomsma, A. (1998). Robustness studies in covariance structure modeling. An overview and a meta-analysis. *Sociological Methods & Research*, 26, 329–367.
- Hord, S. M. (1997). *Professional learning communities. Communities of continuous inquiry and improvement*. Austin: Southwest Educational Development Laboratory.
- Hord, S. M., & Sommers, W. A. (2008). *Leading professional learning communities: voices from research and practice*. Thousand Oaks: Corwin.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.
- Jackson, D., & Tasker, R. (2002). *Professional learning communities*. Cranfield: National College for School Leadership.
- Jak, S., & Jørgensen, T. D. (2017). Relating measurement invariance, cross-level invariance, and multilevel reliability. *Frontiers in Psychology*, 8(1640), 1–9.
- Keller, U., Reith, F., & Metje, B. (2017). Empirische Forschungsmethoden. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion. Inhaltfelder, Forschungsperspektiven und methodische Zugänge* (S. 27–63). Wiesbaden: Springer.
- Klein, E. D. (2016). Zwillinge oder entfernte Verwandte? Strategien der Schulleitung in deprivierter Lage in Deutschland und den USA. *Zeitschrift für Bildungsforschung*, 6(1), 5–23.
- Kruse, S. D., Louis, K. S., & Bryk, A. S. (1995). An emerging framework for analyzing school professional community. In K. S. Louis & S. D. Kruse (Hrsg.), *Professionalism and community: perspectives on reforming urban schools* (S. 23–44). Thousand Oaks: Corwin.
- Lavié, J. M. (2006). Academic discourses on school-based teacher collaboration: revisiting the arguments. *Educational Administration Quarterly*, 42(5), 773–805.
- Leithwood, K., & Sun, J. (2012). The nature and effects of transformational school leadership: a meta-analytic review of unpublished research. *Educational Administration Quarterly*, 48(3), 387–423.
- Li, L., Hallinger, P., & Walker, A. (2016). Exploring the mediating effects of trust on principal leadership and teacher professional learning in Hong Kong primary schools. *Educational Management Administration & Leadership*, 44(1), 20–42.
- Little, T. D. (2013). *Longitudinal structural equation modeling. Methodology in the social sciences*. New York: Guilford.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: exploring the question and weighing the merits. *Structural Equation Modeling*, 9, 151–173.
- Lomos, C., Hofman, R. H., & Bosker, R. J. (2011a). Professional communities and student achievement—a meta-analysis. *School Effectiveness and School Improvement*, 22(2), 121–148.
- Lomos, C., Hofman, R. H., & Bosker, R. J. (2011b). The relationship between departments as professional communities and student achievement in secondary schools. *Teaching and Teacher Education*, 27(4), 722–731.
- Louis, K. S., Dretzke, B., & Wahlstrom, K. (2010). How does leadership affect student achievement? Results from a national US survey. *School Effectiveness and School Improvement*, 21(3), 315–336.
- Louis, K. S., Murphy, J., & Smylie, M. (2016). Caring leadership in schools: findings from exploratory analyses. *Educational Administration Quarterly*, 52(2), 310–348.
- Lüders, J. (Hrsg.). (2007). *Fachkulturforchung in der Schule*. Farmington Hills: Barbara Budrich.
- Lüdtke, O., Marsh, H., Robitzsch, A., Trautwein, U., Asparouhov, T., & Muthén, B. (2008). The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. *Psychological Methods*, 13, 203–229.
- Lüdtke, O., Trautwein, U., Kunter, M., & Baumert, J. (2006). Analyse von Lernumwelten: Ansätze zur Bestimmung der Reliabilität und Übereinstimmung von Schülerwahrnehmungen. *Zeitschrift für Pädagogische Psychologie*, 20(1/2), 85–96.
- Marsh, H. W., Lüdtke, O., Robitzsch, A., Trautwein, U., Asparouhov, T., Muthén, B., & Nagengast, B. (2009). Doubly-latent models of school contextual effects: integrating multilevel and structural equation approaches to control measurement and sampling error. *Multivariate Behavioral Research*, 44(6), 764–802.
- McLaughlin, M., & Talbert, J. E. (2007). Building professional learning communities in high schools: challenges and promising practices. In L. Stoll & K. S. Louis (Hrsg.), *Professional learning communities: divergence, depth, and dilemmas* (S. 151–165). Berkshire: McGraw-Hill Open University Press.
- Mitchell, C., & Sackney, L. (2007). Extending the learning community: a broader perspective embedded in policy. In L. Stoll & K. S. Louis (Hrsg.), *Professional learning communities: divergence, depth and dilemmas* (S. 30–44). Berkshire: McGraw-Hill Open University Press.

- Mitchell, C., & Sackney, L. (2016). School improvement in high-capacity schools: educational leadership and living-systems ontology. *Educational Management Administration & Leadership*, 44(5), 853–868.
- Moller, G., & Pankake, A. (2006). *Lead with me: a principal's guide to teacher leadership*. New York: Taylor & Francis.
- Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(4), 599–620.
- Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8. Aufl.). Los Angeles: Muthén & Muthén.
- OECD (2010). Education at a glance. <https://www.oecd.org/education/skills-beyond-school/45926093.pdf>. Zugegriffen: 30. Dez. 2016.
- Olivier, D. F., Hipp, K. K., & Huffman, J. B. (2010). Assessing and analyzing schools as professional learning communities. In K. K. Hipp & J. B. Huffman (Hrsg.), *Demystifying professional learning communities: school leadership at its best*. Lanham: Rowman & Littlefield.
- Pröbstel, C. H., & Soltan, A. (2012). Wieso Lehrkräfte (nicht) kooperieren – Die Bedeutung „personaler Faktoren“ in der Zusammenarbeit am Arbeitsplatz Schule. In E. Baum, T.-S. Idel & H. Ullrich (Hrsg.), *Kollegialität und Kooperation in der Schule. Theoretische Konzepte und empirische Befunde* (S. 55–76). Wiesbaden: Springer VS.
- Röder, L. (2017). *Kollegiale Teamarbeit an Berufsbildenden Schulen in Hessen: Empirische Befunde zu Implementierung und Qualität*. Bern: Peter Lang.
- Rone, B. C. (2009). *The impact of the data team structure on collaborative teams and student achievement (Nicht veröffentlichte Dissertation)*. Lindenwood University
- Salleh, H., Goh, J. W. P., Chua, C. S. K., & Wang, L. Y. (2017). A research agenda for professional learning communities: moving forward. *Professional Development in Education*, 43(1), 72–86.
- Scribner, J. P., Hager, D. R., & Warne, T. R. (2002). The paradox of professional community: tales from two high schools. *Educational Administration Quarterly*, 38(1), 45–76.
- Sigurðardóttir, A. K. (2010). Professional learning community in relation to school effectiveness. *Scandinavian Journal of Educational Research*, 54(5), 395–412.
- Slegers, P., Den Brok, P., Verbiest, E., Moolenaar, N., Nienke, M., & Daly, A. J. (2013). Toward conceptual clarity. A multidimensional, multilevel model of professional learning communities in Dutch elementary schools. *The Elementary School Journal*, 114(1), 118–137.
- Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *J. Educ. Behav. Statist.*, 41, 481–520.
- Steinert, B., Gerecht, M., Klieme, E., & Döbrich, P. (2003). *Skalen zur Schulqualität: Dokumentation der Erhebungsinstrumente*. Frankfurt a.M.: Gesellschaft zur Förderung Pädagogischer Forschung.
- Stoll, L., Bolam, R., McMahon, A., Wallace, M., & Thomas, S. (2006). Professional learning communities: a review of the literature. *Journal of Educational Change*, 7(4), 221–258.
- Supovitz, J., Sirinides, P., & May, H. (2009). How principals and peers influence teaching and learning. *Educational Administration Quarterly*, 46(1), 31–56.
- Tenberg, R. (2017). Zum Stand der Lehrpersonenkooperation an berufsbildenden Schulen. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 113(2), 179–201.
- Thiel, F., Brauckmann, S., & van Ackeren, I. (2018). Editorial zum Schwerpunktthema: Datenbasiertes Schulleitungshandeln. *Die Deutsche Schule*, 110(1), 5–9.
- Thompson, S. C., Gregg, L., & Niska, J. M. (2004). Professional learning communities, leadership, and student learning. *RMLE Online*, 28(1), 1–15.
- Valckx, J., Devos, G., & Vanderlinde, R. (2018). Exploring the relationship between professional learning community characteristics in departments, teachers' professional development, and leadership. *Pedagogische Studien*, 95, 34–55.
- Van Ackeren, I., Block, R., Kullmann, H., Sprütten, F., & Klemm, K. (2008). Schulkultur als Kontext naturwissenschaftlichen Lernens. Allgemeine und fachspezifische explorative Analysen. *Zeitschrift für Pädagogik*, 54(3), 341–360.
- Van Houtte, M. (2005). Climate or culture? A plea for conceptual clarity in school effectiveness research. *School Effectiveness and School Improvement*, 16(1), 71–89.
- Vanblaere, B., & Devos, G. (2016). Relating school leadership to perceived professional learning community characteristics: a multilevel analysis. *Teaching and Teacher Education*, 57, 26–38.
- Vanblaere, B., & Devos, G. (2018). Relating school leadership to perceived professional learning community characteristics: a multilevel analysis. *Educational Administration Quarterly*, 54(1), 85–114.
- Vangrieken, K., Dochy, F., Raes, E., & Kynndt, E. (2015). Teacher collaboration: a systematic review. *Educational Research Review*, 15, 17–40.

- Vangrieken, K., Meredith, C., Packer, T., & Kyndt, E. (2017). Teacher communities as a context for professional development: A systematic review. *Teaching and Teacher Education*, *61*, 47–59.
- Wahlstrom, K.L., & Louis, K.S. (2008). How teachers experience principal leadership: the roles of community, trust, efficacy, and shared responsibility. *Educational Administration Quarterly*, *44*(4), 458–495.
- Wang, T. (2015). Contrived collegiality versus genuine collegiality: demystifying professional learning communities in Chinese schools. *Compare: A Journal of Comparative and International Education*, *45*(6), 908–930.
- Warwas, J., & Helm, C. (2018). Professional learning communities among vocational school teachers—profiles and relations with instructional quality. *Teaching and Teacher Education*, *73*, 43–55.
- Webs, T., & Holtappels, H.G. (2018). School conditions of different forms of teacher collaboration and their effects on instructional development in schools facing challenging circumstance. *Journal of Professional Capital and Community*, *3*(1), 39–58.
- Wells, C., & Feun, L. (2007). Implementation of learning community principles: a study of six high schools. *NASSP Bulletin*, *91*(2), 141–160.
- Wissinger, J. (2014). Schulleitungshandeln und Schulentwicklung. In H.G. Holtappels (Hrsg.), *Schulentwicklung und Schulwirksamkeit als Forschungsfeld: Theorieansätze und Forschungserkenntnisse zum schulischen Wandel* (S. 123–140). Münster: Waxmann.
- Zhang, J., Yuan, R., & Yu, S. (2017). What impedes the development of professional learning communities in China? Perceptions from leaders and frontline teachers in three schools in Shanghai. *Educational Management Administration & Leadership*, *45*(2), 219–237.
- Zheng, X., Yin, H., Liu, Y., & Ke, Z. (2016). Effects of leadership practices on professional learning communities: the mediating role of trust in colleagues. *Asia Pacific Education Review*, *17*(3), 521–532.
- Zitzmann, S., Lüdtke, O., & Robitzsch, A. (2015). A Bayesian approach to more stable estimates of group-level effects in contextual studies. *Multivariate Behavioral Research*, *50*(6), 688–705.
- Zitzmann, S., Lüdtke, O., Robitzsch, A., & Marsh, H.W. (2016). A Bayesian approach for estimating multilevel latent contextual models. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(5), 661–679.