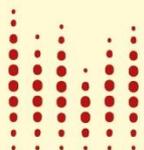


Julia Larissa Maier, Tosca Panetta, Jana Philipp,
Evelin Ruth-Herbein & Benjamin Fauth

Unterrichtsfeedbackbogen Tiefenstrukturen

Präpilotierung und Validierung:
Erste Ergebnisse der wissenschaftlichen Begleitung



IBBW
Institut für Bildungsanalysen
Baden-Württemberg

UFB 
IBBW –
Unterrichtsfeedbackbogen



Baden-Württemberg

Oktober 2021

INHALTSVERZEICHNIS

| | |
|---|----|
| Kurzzusammenfassung | 2 |
| 1. Hintergrund | 2 |
| 2. Präpilotierung | 4 |
| 2.1 Fragestellungen | 4 |
| 2.2 Ablauf | 4 |
| 2.3 Ergebnisse | 5 |
| 3. Validierung | 10 |
| 3.1 Fragestellung | 10 |
| 3.2 Ablauf | 10 |
| 3.3 Ergebnisse | 11 |
| 4. Implikationen | 13 |
| 4.1 Implementation Qualifizierungsveranstaltung | 13 |
| 4.2 Implementation Instrumentarium | 14 |
| 4.3 Austausch und inhaltliche Schärfung | 14 |
| 4.4 Psychometrische Güte | 14 |
| 5. Ausblick | 15 |
| Literatur | 16 |
| Anhang | 17 |
| Impressum | 18 |

Kurzzusammenfassung:

Im vorliegenden Bericht werden die ersten Ergebnisse der wissenschaftlichen Begleitung des Projekts „Qualitätsentwicklung durch Unterrichtsbeobachtung und Feedback“ dargestellt.

Die Studien liefern erste Hinweise darauf, dass

- die Qualifizierungsveranstaltung und das Instrumentarium von den Teilnehmenden positiv eingeschätzt werden.
- unabhängige Beobachterinnen und Beobachter mithilfe des UFB zu übereinstimmenden Einschätzungen der Unterrichtsqualität kommen (*Reliabilität*).
- der UFB die Merkmale der unterrichtlichen Tiefenstrukturen erfasst, die er erfassen soll (*Validität*).

Die Erkenntnisse der wissenschaftlichen Begleitung tragen zur kontinuierlichen Weiterentwicklung des Projekts bei.

1. Hintergrund

Welche Unterrichtsmerkmale zeichnen guten und wirksamen Unterricht aus und wie können diese Merkmale beobachtet und Lehrkräften rückgemeldet werden? Dies waren die zentralen Fragen bei der Entwicklung des Unterrichtsfeedbackbogens Tiefenstrukturen.

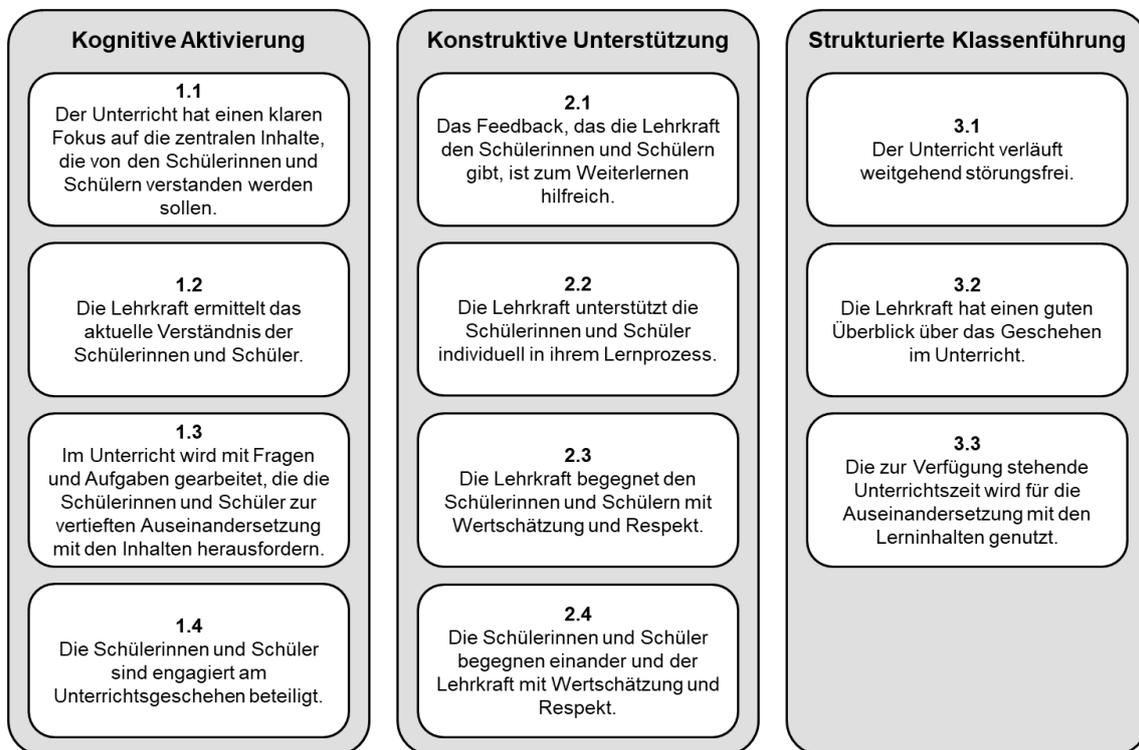
In der Unterrichtsqualitätsforschung hat sich in vielen Studien gezeigt, dass drei Tiefenstrukturen von Unterrichtsqualität bei der Beschreibung guten und wirksamen Unterrichts eine bedeutsame Rolle spielen: kognitive Aktivierung, konstruktive Unterstützung und strukturierte Klassenführung. Diese werden als die Basisdimensionen von Unterrichtsqualität bezeichnet. In empirischen Studien konnte wiederholt gezeigt werden, dass sich die Basisdimensionen positiv auf die Lernentwicklung und auf die motivationale Entwicklung von Schülerinnen und Schülern auswirken (z. B. Fauth et al., 2014; Kunter & Voss, 2011; Lipowsky et al., 2009). Anders als die unterrichtlichen Sichtstrukturen, zu denen die Organisations- und Sozialformen sowie die Unterrichtsmethoden gehören, sind diese Tiefenstrukturen der Unterrichtsqualität allerdings nicht ohne geeignete Beobachtungsinstrumente erfassbar. Hier setzt das Projekt „Qualitätsentwicklung durch Unterrichtsbeobachtung und Feedback“ an. Im Rahmen des Projekts wurde am Institut für Bildungsanalysen Baden-Württemberg der Unterrichtsfeedbackbogen Tiefenstrukturen (UFB; Fauth et al., 2021) auf der Grundlage wissenschaftlicher Erkenntnisse konzipiert und praktisch erprobt. Der UFB soll die Tiefenstrukturen des Unterrichts greifbar, beschreibbar und damit für die Praxis nutzbar machen. So soll ein gemeinsames Verständnis von qualitativem Unterricht hergestellt werden. Der UFB soll es Lehrkräften ermöglichen, detaillierte und lernförderliche Rückmeldungen zu ihrem Unterricht zu geben und zu erhalten. Dies kann beispielsweise im Rahmen der Aus- und Fortbildung, der kollegialen Hospitation oder auch der Selbstreflexion geschehen. Mit dem UFB soll

bewusst nicht die Qualität des Unterrichts in seiner gesamten Breite beschrieben werden. Stattdessen wird auf ausgewählte, aber besonders relevante Merkmale der unterrichtlichen Tiefenstrukturen fokussiert. In Abbildung 1 sind die mit dem UFB erfassbaren Aspekte der Unterrichtsqualität dargestellt. Weiterführende Informationen zum Projekt sowie die Möglichkeit zum Download des UFB und des Beobachtungsmaterials finden Sie auf der [Homepage des IBBW](#).

Der UFB sowie die begleitenden Unterstützungssysteme (Beobachtungsmaterial, Qualifizierungsveranstaltung) wurden und werden weiterhin in mehreren Feedback- und Erprobungsphasen schrittweise angepasst und weiterentwickelt. Auf diese Weise entsteht ein wissenschaftlich fundiertes und optimiertes Instrumentarium samt Unterstützungssystemen. Gegenstand des vorliegenden Ergebnisberichts sind die Phasen der Präpilotierung (September bis November 2020) und Validierung (November 2020 bis Februar 2021) des UFB.

Abbildung 1

Aspekte der Unterrichtsqualität im Unterrichtsfeedbackbogen Tiefenstrukturen



2. Präpilotierung

2.1 FRAGESTELLUNGEN

Die Präpilotierungsstudie hatte eine erste Erprobung des Instrumentariums, bestehend aus UFB und Beobachtungsmanual, sowie der Qualifizierungsveranstaltung zum Ziel. Diese Projektphase wurde außerdem für einen intensiven Austausch mit den teilnehmenden Lehrkräften genutzt, auf dessen Basis Anpassungen zur inhaltlichen Schärfung und Hinweise für die praktische Anwendung des Instrumentariums abgeleitet wurden.

Neben Fragen zur Implementation wurde untersucht, inwieweit mithilfe des UFB ein gemeinsames Verständnis und eine übereinstimmende Einschätzung der Unterrichtsqualität erreicht werden kann.

Fragestellungen:

1. Wie wird die Qualifizierungsveranstaltung von den Teilnehmenden eingeschätzt?
2. Wie wird das Instrumentarium von den Teilnehmenden eingeschätzt?
3. Wie gut stimmen unabhängige Beobachterinnen und Beobachter nach Durchlaufen der Qualifizierungsveranstaltung in ihren Urteilen überein?

2.2 ABLAUF

Die Präpilotierungsstudie fand im Zeitraum von September bis November 2020 statt. An der Studie nahmen $N = 10$ Personen teil. Alle teilnehmenden Personen sind ausgebildete Lehrkräfte, welche in der Lehrkräfteausbildung und -fortbildung für das Fach Mathematik tätig sind (jeweils die Hälfte für das Gymnasium und die Sekundarstufe 1).

Die Teilnehmenden durchliefen eine Online-Qualifizierungsveranstaltung, welche vom Entwicklungsteam des IBBW und des ZSL durchgeführt wurden. Sie bestand aus synchronen und asynchronen Phasen und umfasste mehrere inhaltliche Bausteine (Inputphasen, Trainingsphasen sowie Austauschphasen zur gemeinsamen Reflexion und Diskussion).

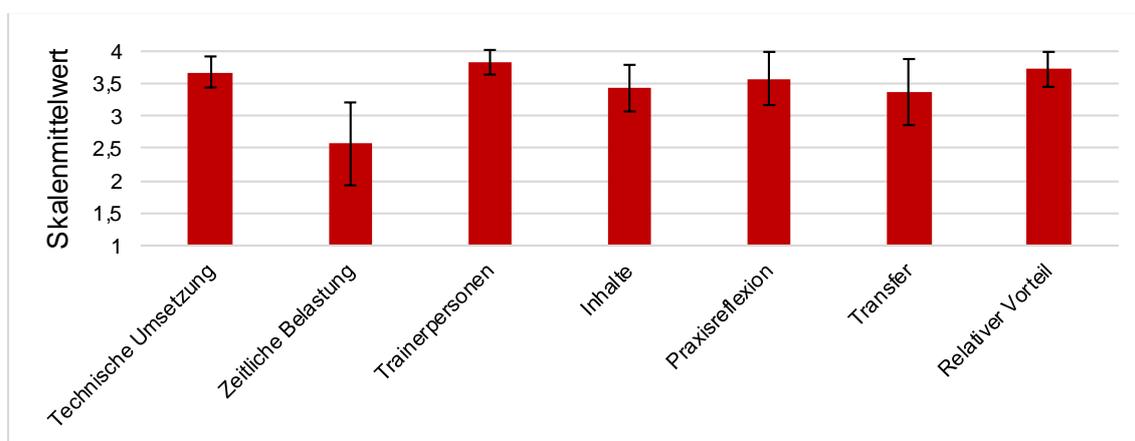
Vor, zwischen und nach den Bausteinen fanden Online-Befragungen statt. In Abbildung 2 sind die einzelnen inhaltlichen Bausteine sowie die Messzeitpunkte dargestellt. Die gewonnenen Daten der Zwischenbefragungen wurden und werden für die Weiterentwicklung des Instrumentariums und der Qualifizierungsveranstaltung verwendet. Dieser Ergebnisbericht fokussiert auf die Ergebnisse nach Abschluss der Qualifizierungsveranstaltung (Posttest: T5).

Verlauf der Qualifizierung und deren Ergebnissen zufrieden waren ($M = 3.70$, $SD = 0.48$), hat ihnen der reale Face-to-face Kontakt gefehlt ($M = 1.90$, $SD = 0.88$).

Des Weiteren berichten die Teilnehmenden von einer mittleren bis leicht erhöhten zeitlichen Belastung durch die Qualifizierungsveranstaltung. Die Trainerpersonen (Entwicklungsteam des IBBW und ZSL) wurden positiv bewertet. Auf inhaltlicher Ebene wurde die Qualifizierungsveranstaltung ebenfalls positiv eingeschätzt. Hohe Ausprägungen in den erhobenen Skalen zeigen außerdem, dass die Veranstaltung zur Reflexion der eigenen Praxis anregte und Transfermöglichkeiten zur Übertragung der Inhalte in die Praxis bot. Die Teilnehmenden nahmen zudem einen relativen Vorteil der Qualifizierungsveranstaltung im Vergleich zu bestehenden Veranstaltungen wahr.

Abbildung 3

Implementation der Qualifizierungsveranstaltung zu T5 (Posttest)



Anmerkung: Es werden die Mittelwerte der berechneten Skalen abgebildet. Die Fehlerbalken entsprechen den jeweiligen Standardabweichungen. Exakte Werte sind im Anhang 1 dargestellt.

Tabelle 1

Beispielitems zur Implementation der Qualifizierungsveranstaltung

| Skala | Beispielitem |
|----------------------|--|
| Technische Umsetzung | In den synchronen Phasen der Veranstaltung war der Ton gut verständlich. |
| Zeitliche Belastung | Die Arbeitsbelastung durch die Veranstaltung ist für mich zu hoch. |
| Trainerpersonen | Die Referentinnen/Referenten waren kompetent. |
| Inhalte | Die Qualifizierungsveranstaltung war informativ. |
| Praxisreflexion | Mein eigenes Vorgehen beim Beobachten von Unterricht ist mir durch die Veranstaltung bewusster geworden. |
| Transfer | Ich habe in der Veranstaltung etwas gelernt, was ich unmittelbar beim Einsatz des UFB in der Praxis anwenden kann. |
| Relativer Vorteil | Die Qualifizierungsveranstaltung ist eine relevante Ergänzung bestehender Fortbildungsangebote. |

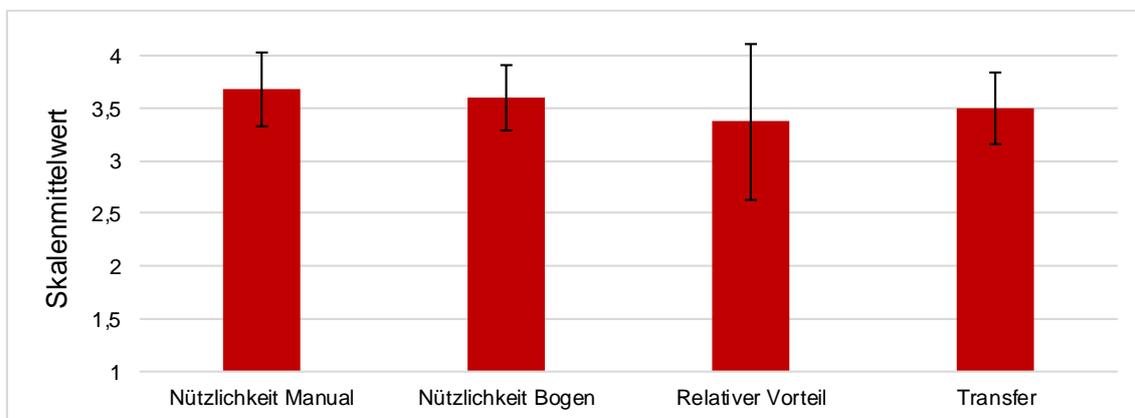
Fragestellung 2: Wie wird das Instrumentarium von den Teilnehmenden eingeschätzt?

Das Instrumentarium, bestehend aus UFB und Beobachtungsmanual, wird von den Teilnehmenden nach Ende der Qualifizierungsveranstaltung positiv eingeschätzt.

Die Teilnehmenden wurden gebeten, dem UFB und dem Beobachtungsmanual eine Schulnote zu geben: Der Mittelwert lag für den UFB bei 1.5, für das Beobachtungsmanual bei 1.4. Darüber hinaus wurden auf einer 4-stufigen Skala von „trifft nicht zu“ bis „trifft völlig zu“ sowohl der UFB selbst als auch das Beobachtungsmanual als nützlich wahrgenommen. Hohe Ausprägungen in den erhobenen Skalen zeigen, dass die Teilnehmenden einen relativen Vorteil des UFB im Vergleich zu bestehenden Beobachtungsinstrumenten erkannten. Des Weiteren nahmen die Teilnehmenden Transfermöglichkeiten zur Übertragung der Inhalte in die Praxis wahr. In Abbildung 4 sind die Skalenmittelwerte sowie Standardabweichungen der erhobenen Aspekte zur Implementation dargestellt. Tabelle 2 zeigt entsprechende Beispielitems.

Abbildung 4

Implementation des Instrumentariums zu T5 (Posttest)



Anmerkung: Es werden die Mittelwerte der berechneten Skalen abgebildet. Die Fehlerbalken entsprechen den jeweiligen Standardabweichungen. Exakte Werte sind im Anhang 2 dargestellt.

Tabelle 2

Beispielitems zur Implementation des Instrumentariums

| Skala | Beispielitem |
|----------------------|---|
| Nützlichkeits Manual | Das Manual ist sehr nützlich für meine Arbeit mit dem Unterrichtsfeedbackbogen. |
| Nützlichkeits Bogen | Der Bogen ist sehr nützlich für mich um Unterricht zu reflektieren. |
| Relativer Vorteil | Der Einsatz des UFB ermöglicht es mir, den Unterricht strukturierter als bisher zu beobachten. |
| Transfer | Den UFB kann ich später unmittelbar zur Beobachtung der Tiefenstrukturen im Unterricht einsetzen. |

Fragestellung 3: Wie gut stimmen unabhängige Beobachterinnen und Beobachter nach Durchlaufen der Qualifizierungsveranstaltung in ihren Urteilen überein?

Nach Ende der Qualifizierungsveranstaltung konnte für alle Items des UFB eine zufriedenstellende Beobachtungsübereinstimmung erreicht werden.

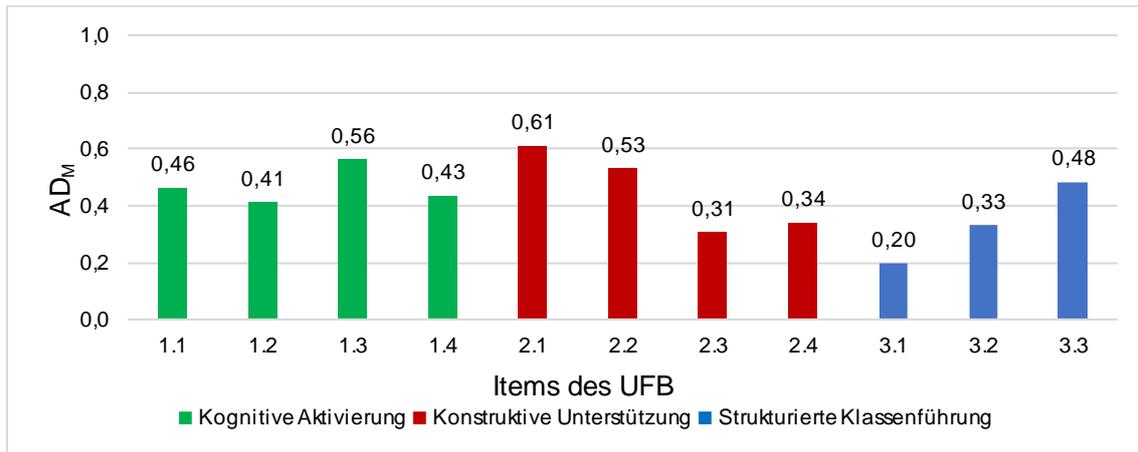
Als Maß für die Beobachtungsübereinstimmung wurde der „Average Deviation Index“ (AD_M ; zu dt.: durchschnittlicher Abweichungsindex; Burke et al., 1999) verwendet. Der AD_M beschreibt die mittlere absolute Abweichung einer Gruppe vom Gruppenmittelwert. Die Werte können in der ursprünglichen Metrik des Items interpretiert werden. Beispielsweise bedeutet ein AD_M von 0.5, dass die Urteilenden im Durchschnitt um einen halben Punkt auf der Antwortskala vom Gruppenmittelwert abweichen. Ein hoher AD_M steht für eine hohe Abweichung der einzelnen Urteilenden untereinander und deutet auf eine geringe Beobachtungsübereinstimmung hin. Ein niedriger AD_M steht für eine niedrige Abweichung der einzelnen Urteilenden untereinander und deutet auf eine hohe Beobachtungsübereinstimmung hin. Demnach sind niedrige Ausprägungen des AD_M erstrebenswert. In der wissenschaftlichen Literatur werden unterschiedliche Grenzwerte diskutiert, ab wann von einer zufriedenstellenden Beobachtungsübereinstimmung gesprochen werden kann. Bei der im UFB verwendeten 4-stufigen Skala (von „trifft nicht zu“ bis „trifft völlig zu“) wäre beispielsweise nach Burke et al. (1999) ein $AD_M < 0.67$ zufriedenstellend, nach Smith-Crowe et al. (2014) wäre ein $AD_M < 0.60$ zufriedenstellend.

Nach Ende der Qualifizierungsveranstaltung schätzten die teilnehmenden Personen zehn Videosequenzen anhand der elf Items des UFB ein. Gemittelt über alle Videos und alle Items zeigte sich ein AD_M von 0.42. Allerdings variierte der AD_M zwischen den einzelnen Items: Das Item 3.1 (Störungen) weist mit einem AD_M von 0.20 den geringsten Wert auf, das Item 2.1 (Qualität des Feedbacks) mit einem AD_M von 0.61 den höchsten Wert. Die exakten AD_M -Werte der Einzelitems sind in Abbildung 5 dargestellt. Diese Unterschiede weisen auf eine unterschiedlich gute Beobachtbarkeit einzelner Unterrichtsqualitätsaspekte hin. Items, die auch in der wissenschaftlichen Literatur als leichter zu beobachten beschrieben werden (vgl. Clausen, 2002; Fauth et al., 2020) zeigen auch hier geringere AD_M -Werte. Schwieriger zu beobachtende Items zeigen höhere AD_M -Werte. Des Weiteren spiegeln diese Unterschiede auch die Diskussionsphasen der Qualifizierungsveranstaltung wider: Items, die aufgrund von Unsicherheiten oder Uneinigkeiten verstärkt mit den Teilnehmenden diskutiert wurden, zeigten auch im Posttest eine geringere Beobachtungsübereinstimmung. Bei diesen Items zeigte sich auch der größte Anpassungsbedarf an Formulierungen auf Item- sowie Indikatorebene.

Insgesamt kann die Beobachtungsübereinstimmung als zufriedenstellend interpretiert werden: Unabhängige Beobachterinnen und Beobachter stimmen nach Durchlaufen der Qualifizierungsveranstaltung weitestgehend in ihren Urteilen überein.

Abbildung 5

Beobachtungsübereinstimmung (mittlerer AD_M) zu T5 (Posttest)

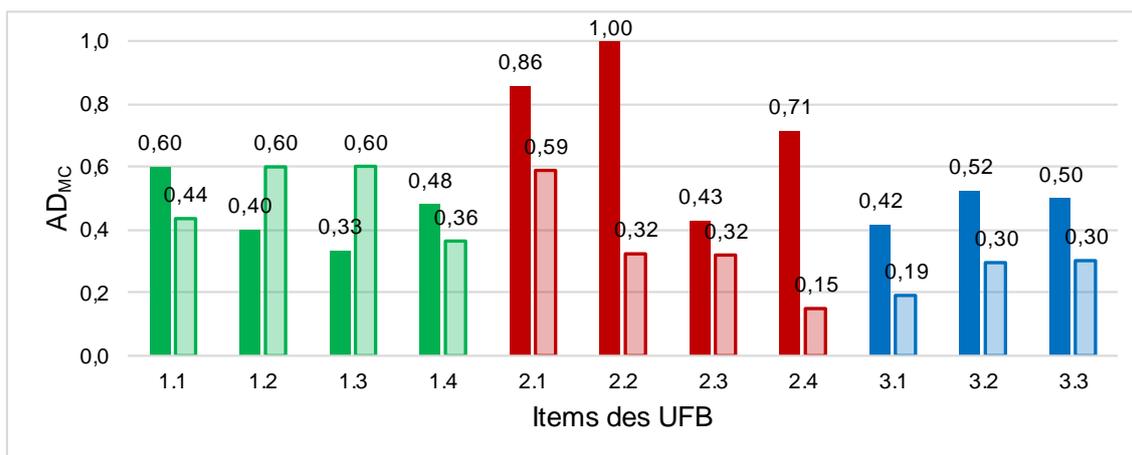


Anmerkung: Die Formulierungen der einzelnen Items sind in Abbildung 1 dargestellt.

Für zwei Messzeitpunkte im Verlauf der Präpilotierung (T3 und T4) lagen zusätzlich zu den Einschätzungen der Teilnehmenden auch Einschätzungen des Entwicklungsteams am IBBW vor (sogenannte Masterratings). Zu T3 wurden drei, zu T5 wurden fünf ca. 5-minütige Videosequenzen eingeschätzt. Für diese zwei Messzeitpunkte konnte dadurch ergänzend zur mittleren Abweichung der Teilnehmenden vom Gruppenmittelwert auch die mittlere Abweichung der Teilnehmenden zu den Masterratings bestimmt werden. In Abbildung 6 ist zu sehen, dass die Einschätzungen der Teilnehmenden in der Mitte der Qualifizierungsveranstaltung (T3) noch stärker von den Masterratings abwichen, als gegen Ende der Qualifizierungsveranstaltung (T4). Für die meisten Items zeigte sich eine positive Entwicklung in der Beobachtungsübereinstimmung zwischen Teilnehmenden und dem Entwicklungsteam am IBBW im Verlauf der Qualifizierungsveranstaltung.

Abbildung 6

Beobachtungsübereinstimmung zwischen Teilnehmenden und Masterratings zu T3 und T4



Anmerkung: Die ausgefüllten, linken Balken zeigen die Beobachtungsübereinstimmungen zu T3. Die halbtransparenten, rechten Balken zeigen die Beobachtungsübereinstimmung zu T4. Die Formulierungen der einzelnen Items sind in Abbildung 1 dargestellt.

3. Validierung

3.1 FRAGESTELLUNGEN

An die Präpilottierungsphase schloss sich die Phase der Validierung an. Diese hatte eine genauere Überprüfung der psychometrischen Güte des Instrumentariums zum Ziel. Neben der Reliabilität stand die Frage der Validität im Fokus. Ein Messinstrument wird als reliabel bezeichnet, wenn es das Merkmal, das es messen soll, genau und zuverlässig misst (Genauigkeit der Einschätzung). Die Validität beschäftigt sich hingegen mit der Frage, ob ein Messinstrument überhaupt das Merkmal misst, welches es messen soll (Gültigkeit der Einschätzung).

Fragestellungen:

1. Reliabilität: Wie gut stimmen unabhängige Beobachterinnen und Beobachter in ihren Urteilen überein?
2. Validität: Erfasst der UFB die Facetten der Unterrichtsqualität, die er erfassen soll?

3.2 ABLAUF

Die Validierungsstudie fand im Zeitraum von November 2020 bis Februar 2021 statt. An der Studie nahmen die Personen teil, welche im Rahmen der Präpilottierung intensiv im Umgang mit dem UFB und dem Beobachtungsmanual geschult wurden ($N = 10$).

Für die Validierungsstudie wurden ca. 45-minütige Unterrichtsvideos der Pythagoras-Studie (Klieme et al., 2009) verwendet. Diese zeigen Mathematikunterricht aus schweizerischen und deutschen Schulklassen der achten bzw. neunten Jahrgangsstufe zum Thema „Einführung in die Satzgruppe des Pythagoras“.

Die Teilnehmenden betrachteten in sechs Blöcken insgesamt 37 Unterrichtsvideos und gaben ihre Einschätzung zur Unterrichtsqualität anhand des UFB ab. In den ersten fünf Blöcken wurden jeweils alle Unterrichtsvideos von allen Teilnehmenden betrachtet und eingeschätzt. Im letzten Block wurden zwölf Unterrichtsvideos randomisiert auf die zehn Teilnehmenden verteilt, sodass jede Person fünf Videos betrachtete und einschätzte. Zwischen den einzelnen Blöcken fanden teilweise digitale Austauschtreffen mit dem Entwicklungsteam des IBBW statt, um Einschätzungen zu reflektieren und mögliche Unsicherheiten zu klären.

Um zu überprüfen, ob der UFB die Facetten der Unterrichtsqualität erfasst, die er erfassen soll (Validität), wurden die Daten der Validierungsstudie mit Daten der Pythagoras-Studie verglichen. Da in beiden Studien dieselben Unterrichtsvideos betrachtet und eingeschätzt wurden, konnten die entsprechenden Einschätzungen verglichen werden. Zur Überprüfung der Validität wurden inhaltlich identische oder ähnliche Aspekte der Unterrichtsqualität zwischen dem UFB-Instrumentarium und dem Instrumentarium der Pythagoras-Studie identifiziert. Im Sinne der konvergenten Validität sollten verschiedene Instrumente, die dieselben Merkmale messen, zu vergleichbaren Ergebnissen führen. Daher wurden für diese Aspekte im Sinne der konvergenten Validität positive Zusammenhänge zwischen den Ratings anhand des UFB und den bereits vorhandenen und validierten Ratings der Pythagoras-Studie erwartet.

3.3 ERGEBNISSE

Fragestellung 1: Wie gut stimmen unabhängige Beobachterinnen und Beobachter in ihren Urteilen überein?

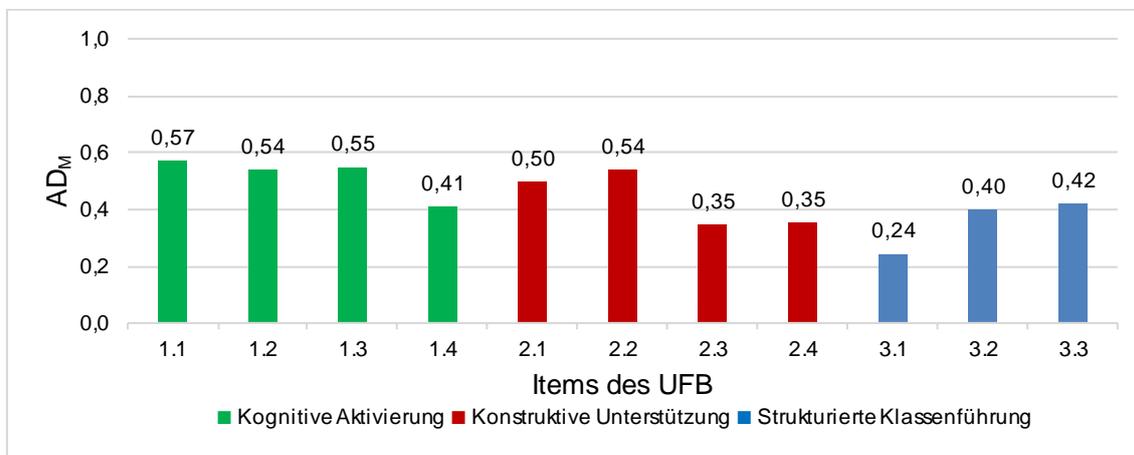
In der Validierungsphase konnte für alle Items des UFB eine zufriedenstellende Beobachtungsübereinstimmung erreicht werden.

Während die Teilnehmenden im Rahmen der Präpilotierung lediglich kurze, ca. 5-minütige Videosequenzen betrachteten, wurden in der Validierung erstmals ganze Unterrichtsstunden anhand des UFB eingeschätzt. Die ersten drei Videos dienten daher der Eingewöhnung in die Einschätzung ganzer Unterrichtsstunden. In diesem Block zeigte sich gemittelt über alle Videos und alle Items ein AD_M von 0.50 (vgl. Posttest der Präpilotierung: $AD_M = 0.42$). Für die folgenden 34 Unterrichtsvideos ergab sich gemittelt über alle Videos und alle Items ein AD_M von 0.42. Alle Werte lagen unter 0.60. Die genauen AD_M -Werte der Einzelitems sind in Abbildung 7 dargestellt. Analog zur Präpilotierung variierte die Beobachtungsübereinstimmung zwischen den Items, was auf die unter-

schiedlich gute Beobachtbarkeit einzelner Unterrichtsqualitätsaspekte hindeutet. Insgesamt ist die Beobachtungsübereinstimmung im Rahmen der Validierung ebenfalls zufriedenstellend.

Abbildung 7

Beobachtungsübereinstimmung (mittlerer AD_M) in der Validierung



Anmerkung: Die Formulierungen der einzelnen Items sind in Abbildung 1 dargestellt.

Fragestellung 2: Erfasst der UFB die Facetten der Unterrichtsqualität, die er erfassen soll?

Signifikante Zusammenhänge mit den Unterrichtsqualitäts-Ratings der Pythagoras-Studie liefern Hinweise auf die Validität des UFB.

Zur Untersuchung der konvergenten Validität wurden die Einschätzungen der Teilnehmenden anhand des UFB mit den bereits vorliegenden Unterrichtsqualitäts-Ratings der Pythagoras-Studie verglichen.

Tabelle 3 zeigt die inhaltlich ähnlichen Aspekte der Unterrichtsqualität zwischen UFB und Pythagoras-Studie. In der Tabelle werden die erwarteten Zusammenhänge sowie die empirisch ermittelten Korrelationen dargestellt. Es finden sich signifikant positive Korrelationen zwischen den UFB-Items und den Unterrichtsqualitäts-Items der Pythagoras-Studie. Die Stärke der Zusammenhänge variiert zwischen den Items, dennoch zeigen sich in fast allen Fällen erwartungskonforme Korrelationen. Dies bedeutet, dass die Teilnehmenden der Validierungsstudie und die Teilnehmenden der Pythagoras-Studie nach der Betrachtung derselben Unterrichtsvideos zu ähnlichen Einschätzungen der Unterrichtsqualität kamen. Die beiden Instrumente zur Erfassung der Unterrichtsqualität erfassen demnach ähnliche Aspekte, was auf konvergente Validität des UFB hindeutet. Die Tatsache, dass die Einschätzungen aus der Pythagoras-Studie durch Personen mit einem anderen fachlichen Hintergrund (studentische Hilfskräfte) sowie in einem anderen Jahrzehnt erfolgten, dürfte diese Hinweise auf konvergente Validität noch verstärken.

Tabelle 3

Abgleich zwischen UFB und Instrumentarium der Pythagoras-Studie

| UFB-Item | Pythagoras-Item | Hypothese | Korrelation |
|-------------------------------------|-------------------------------------|-----------|-------------|
| Kognitive Aktivierung | | | |
| 1.1 Verständnisorientierung | Lernstatus bewusst machen | . | .09 |
| | Klarheit | + | .08 |
| 1.2 Ermittlung Denkweisen | Exploration des Vorwissens | + | .22 |
| | Exploration von Denkweisen | + | .63** |
| 1.3 Herausforderung | Herausfordernde Probleme | ++ | .30* |
| 1.4 Engagement der SuS | Engagement der SuS | ++ | .47** |
| Konstruktive Unterstützung | | | |
| 2.1 Qualität des Feedbacks | Sachlich-konstruktive Rückmeldungen | + | .55** |
| | Positive Fehlerkultur | + | .40* |
| 2.2 Individuelle Unterstützung | Individualisierung | ++ | .40* |
| 2.3 Wertschätzung und Respekt | Anerkennung Lehrperson | ++ | .55** |
| 2.4 Klassenklima | Lerngemeinschaft | + | .32* |
| | Anerkennung SuS | + | .46** |
| Strukturierte Klassenführung | | | |
| 3.1 Störungen | Disziplinprobleme | + | .78** |
| 3.2 Monitoring | Classroom Management | + | .55** |
| 3.3 Zeitnutzung | Classroom Management | + | .45** |

Anmerkungen: . = keine signifikante Korrelation erwartet; + = schwache signifikante Korrelation erwartet; ++ = moderate signifikante Korrelation erwartet; * $p < .05$; ** $p < .01$.

4. Implikationen

Die Präpilotierungs- und die Validierungsstudie zum UFB hatten eine erste Erprobung des Instrumentariums, bestehend aus UFB und Beobachtungsmanual, sowie der Qualifizierungsveranstaltung zum Ziel. Davon abgeleitet ergaben sich Anpassungen und Weiterentwicklungen auf verschiedenen Ebenen des Projekts.

4.1 IMPLEMENTATION QUALIFIZIERUNGSVERANSTALTUNG

Die Ergebnisse der Präpilotierungsstudie zeigen, dass die Umsetzung und die Inhalte der Qualifizierungsveranstaltung in weiten Teilen positiv eingeschätzt wird. Das rein virtuelle Format, bestehend aus synchronen und asynchronen Phasen, erwies sich als gut geeignet für eine intensive Schulung zur Anwendung des UFB. Es konnten Reflexions-

und Transferprozesse angestoßen werden und die Teilnehmenden erkannten einen relativen Vorteil im Vergleich zu bestehenden Veranstaltungen. Obwohl der zeitliche Aufwand als tendenziell hoch eingeschätzt wurde, wurde in offenen Rückmeldungen immer wieder die Relevanz des intensiven Austauschs, bestehend aus Reflexion und Diskussion, hervorgehoben. Die kritische Einschätzung des zeitlichen Aufwands ist daher vermutlich nicht auf die Inhalte der Qualifizierungsveranstaltung zurückzuführen, sondern auf den großen Zeitaufwand im Rahmen der wissenschaftlichen Begleitung (Fragebögen und Videoratings). Diese wird für die Teilnehmenden im weiteren Projektverlauf eine weitaus geringere Rolle einnehmen. Die differenzierten Rückmeldungen zur Implementation der Qualifizierungsveranstaltung tragen einen wesentlichen Teil zur Weiterentwicklung und Optimierung derselben bei.

4.2 IMPLEMENTATION INSTRUMENTARIUM

Neben der Qualifizierungsveranstaltung wird auch das Instrumentarium, bestehend aus UFB und Beobachtungsmanual, positiv eingeschätzt. Neben der wahrgenommenen Nützlichkeit und der Übertragbarkeit der Inhalte in die Praxis erkennen die Teilnehmenden auch einen relativen Vorteil des Instrumentariums im Vergleich zu bestehenden Beobachtungsinstrumenten. Diese Rückmeldungen der Teilnehmenden zeigen, dass das neu entwickelte Instrumentarium mit seiner Kompaktheit, Stringenz und wissenschaftlichen Fundierung eine Lücke schließt.

4.3 AUSTAUSCH UND INHALTLICHE SCHÄRFUNG

Besonders gewinnbringend in dieser Projektphase war der intensive Austausch mit Teilnehmenden aus der Unterrichtspraxis und der Aus- und Fortbildung. Auf diese Weise konnte der auf wissenschaftlicher Grundlage konzipierte UFB auf seine Praxistauglichkeit geprüft und entsprechend angepasst werden. Die Präpilotierung mündete in eine inhaltliche Schärfung auf Basis der diskutierten Aspekte. Dies betrifft sowohl die Ebene der Items im UFB selbst, als auch die Ebene der Indikatoren im Beobachtungsmanual. Vieldiskutierte Aspekte, wie beispielsweise die Frage der fachlichen Korrektheit unterrichteter Inhalte, die fachliche Expertise der beobachtenden Person und die Rolle der Fachspezifität in der Unterrichtsqualität werden im weiteren Projektverlauf verstärkt in den Blick genommen.

4.4 PSYCHOMETRISCHE GÜTE

Sowohl für kurze Unterrichtssequenzen (im Rahmen der Präpilotierung) als auch für ganze Unterrichtsstunden (im Rahmen der Validierung) zeigt sich eine zufriedenstellende Beobachtungsübereinstimmung, wobei manche Items leichter und manche schwieriger beobachtbar scheinen. Dies deutet auf einen besonderen Schulungsbedarf bei bestimmten Items hin. Nach Abschluss der Qualifizierungsveranstaltung gelangen die Teilnehmenden der Studien bei der Betrachtung desselben Unterrichts zu ähnlichen Einschätzungen. Demnach leistet der UFB einen entscheidenden Beitrag zu einem ge-

meinsamen Verständnis und einer übereinstimmenden Wahrnehmung zentraler Unterrichtsqualitätsaspekte. Zusätzlich konnten im Rahmen der Validierungsstudie Hinweise auf konvergente Validität gewonnen werden. Dies bedeutet, dass der UFB die Merkmale der Unterrichtsqualität erfasst, die er erfassen soll.

Zur Einschätzung der Beobachtungsübereinstimmung verwendeten wir den AD_M als einfach zu interpretierendes Maß. Im Rahmen der wissenschaftlichen Begleitung des Projekts ist dieses Vorgehen wichtig, um die berechneten Werte für die Beobachtungsübereinstimmung einordnen und interpretieren zu können. Für den praktischen Einsatz des UFB ist die klare Definition eines Grenzwertes für die Beobachtungsübereinstimmung allerdings nicht zielführend. Im Rahmen der hier beschriebenen Studien konnte eine gute Beobachtungsübereinstimmung erreicht werden. Beim praktischen Einsatz des UFB kann es jedoch durchaus vorkommen, dass unabhängige Beobachter/innen zu unterschiedlichen Einschätzungen derselben Unterrichtsstunde kommen. Es ist wichtig, sich dieser Unterschiede in den Beobachtungen bewusst zu sein. In allen Kontexten, in denen Unterrichtsbeobachtungen genutzt werden, um über Unterricht ins Gespräch zu kommen, bieten gerade diese Unterschiede in den Beobachtungen wertvolle Ausgangspunkte für die gemeinsame Reflexion von Unterricht.

5. Ausblick

Der UFB wird im Rahmen der wissenschaftlichen Begleitung in mehreren aufeinander aufbauenden Begleitstudien unter wissenschaftlichen Standards erprobt. Die erste Durchführung der Qualifizierungsveranstaltung erfolgte daher unter stark standardisierten Bedingungen. In weiterführenden Projektschritten soll der UFB nun unter zunehmend realeren Praxisbedingungen eingesetzt werden. So fungieren die in der Präpilotierung geschulten Teilnehmenden zukünftig als Multiplikatorinnen und Multiplikatoren und schulen im Rahmen der Pilotierung weitere Lehrkräfte im Umgang mit dem UFB. In diesem Schritt werden auch weitere Unterrichtsfächer in die Erprobung des UFB einbezogen. Während sich die Präpilotierungs- und Validierungsphasen auf den Mathematikunterricht fokussierten, nehmen an der Pilotierung sowohl Mathematik- als auch Deutschlehrkräfte teil. Im weiteren Projektverlauf werden fachspezifische Fragen zunehmend an Bedeutung gewinnen, wenn nach und nach fachspezifische Anregungen für Unterrichtsbeobachtung und Unterrichtsentwicklung ausgearbeitet werden. Auch dieser Prozess wird wissenschaftlich begleitet und evaluiert werden.

Literatur

Burke, M. J., Finkelstein, L. M., & Dusig, M. S. (1999). On average deviation indices for estimating interrater agreement. *Organizational Research Methods*, 2(1), 49–68.

Clausen, M. (2002). *Unterrichtsqualität: Eine Frage der Perspektive?*. Waxmann.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1–9.

Fauth, B., Göllner, R., Lenske, G., Praetorius, A. K., & Wagner, W. (2020). Who sees what? Conceptual considerations on the measurement of teaching quality from different perspectives. *Zeitschrift für Pädagogik*, 66, 138–155.

Fauth, B., Herbein, E., & Maier, J. L. (2021). *Beobachtungsmニュアル zum Unterrichtsfeedbackbogen Tiefenstrukturen*. Institut für Bildungsanalysen Baden-Württemberg.

Grünkorn, J., Klieme, E., Praetorius, A.-K., & Schreyer, P. (2020). *Mathematikunterricht im internationalen Vergleich. Ergebnisse aus der TALIS-Videostudie Deutschland*. DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janík & T. Seidel (Hrsg.), *The power of video studies in investigating teaching and learning in the classroom* (S. 137–160). Waxmann.

Kunter, M., & Voss, T. (2011). Das Modell der Unterrichtsqualität in COACTIV: Eine multikriteriale Analyse. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Hrsg.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (S. 85–113). Waxmann.

Lipowsky, F., Rakoczy, K., Pauli, C., Drollinger-Vetter, B., Klieme, E., & Reusser, K. (2009). Quality of geometry instruction and its short-term impact on students' understanding of the Pythagorean Theorem. *Learning and Instruction*, 19(6), 527–537.

Smith-Crowe, K., Burke, M. J., Cohen, A., & Doveh, E. (2014). Statistical significance criteria for the rWG and average deviation interrater agreement indices. *Journal of Applied Psychology*, 99, 239–261.

Anhang

ANHANG 1: IMPLEMENTATION DER QUALIFIZIERUNGSVERANSTALTUNG ZU T5 (POSTTEST DER PRÄPILOTIERUNG)

| Skala | T5 (Posttest) | | |
|----------------------|---------------|----------|-----------|
| | Itemanzahl | <i>M</i> | <i>SD</i> |
| Technische Umsetzung | 12 | 3.68 | 0.24 |
| Zeitliche Belastung | 6 | 2.57 | 0.64 |
| Trainerpersonen | 6 | 3.83 | 0.19 |
| Inhalte | 6 | 3.43 | 0.36 |
| Praxisreflexion | 4 | 3.58 | 0.41 |
| Transfer | 3 | 3.37 | 0.51 |
| Relativer Vorteil | 5 | 3.72 | 0.27 |

Anmerkungen: *N* = 10; *M* = Skalenmittelwert; *SD* = Standardabweichung.

ANHANG 2: IMPLEMENTATION DES INSTRUMENTARIUMS ZU T5 (POSTTEST DER PRÄPILOTIERUNG)

| Skala | T5 (Posttest) | | |
|---------------------|---------------|----------|-----------|
| | Itemanzahl | <i>M</i> | <i>SD</i> |
| Nützlichkeit Manual | 6 | 3.68 | 0.35 |
| Nützlichkeit Bogen | 5 | 3.60 | 0.31 |
| Relativer Vorteil | 3 | 3.37 | 0.74 |
| Transfer | 8 | 3.50 | 0.34 |

Anmerkungen: *N* = 10; *M* = Skalenmittelwert; *SD* = Standardabweichung.

IMPRESSUM

Oktober 2021

Herausgeber:

Land Baden-Württemberg,
vertreten durch das Institut für Bildungsanalysen Baden-Württemberg (IBBW)
Heilbronner Straße 172
70191 Stuttgart
0711 6642-0
poststelle@ibbw.kv.bwl.de
Vertretungsberechtigter: Direktor Dr. Günter Klein

Autor/Autorinnen:

Julia Larissa Maier, Tosca Panetta, Jana Philipp, Dr. Evelin Ruth-Herbein,
Prof. Dr. Benjamin Fauth
Institut für Bildungsanalysen Baden-Württemberg (IBBW), Stuttgart
Kontakt: unterrichtsfeedbackbogen@ibbw.kv.bwl.de

Verantwortlich im Sinne des Presserechts (RStV):

Dr. Günter Klein
Institut für Bildungsanalysen Baden-Württemberg (IBBW)
Heilbronner Straße 172
70191 Stuttgart

Vertrieb:

Institut für Bildungsanalysen Baden-Württemberg (IBBW)
<https://ibbw-bw.de/Lde/Startseite/Empirische-Bildungsforschung/unterrichtsfeedbackbogen>

Urheberrecht:

Inhalte dieses Heftes dürfen für unterrichtliche Zwecke in den Schulen und Hochschulen des Landes Baden-Württemberg vervielfältigt werden. Jede darüber hinausgehende fotomechanische oder anderweitig technisch mögliche Reproduktion ist nur mit Genehmigung des Herausgebers möglich. Soweit die vorliegende Publikation Nachdrucke enthält, wurden dafür nach bestem Wissen und Gewissen Lizenzen eingeholt. Die Urheberrechte der Copyrightinhaber werden ausdrücklich anerkannt. Sollten dennoch in einzelnen Fällen Urheberrechte nicht berücksichtigt worden sein, wenden Sie sich bitte an den Herausgeber. Bei weiteren Vervielfältigungen müssen die Rechte der Urheber beachtet bzw. muss deren Genehmigung eingeholt werden.

© Institut für Bildungsanalysen Baden-Württemberg (IBBW), Stuttgart 2021



IBBW

Institut für Bildungsanalysen
Baden-Württemberg

UFB

IBBW –
Unterrichtsfeedbackbogen



Baden-Württemberg