

# MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Volksschule ohne Noten?  
Präferenz für Verbalbeurteilung versus Ziffernnoten und  
der Zusammenhang mit *Offenheit*  
sowie die Ungenauigkeit von Schulnoten

verfasst von / submitted by

Philipp Wurmitzer, BSc.

Angestrebter akademischer Grad /  
in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt /  
degree programme code as it appears on  
the student record sheet:

A 066 840

Studienrichtung lt. Studienblatt /  
degree programme as it appears on  
the student record sheet:

Masterstudium Psychologie UG2002

Betreut von / Supervisor:

Univ.-Prof. Mag. DDr. Christiane Spiel



## **Danksagung**

Ich möchte mich bei Frau Mag. DDr. Christiane Spiel für die Unterstützung bei der Themensuche, der begleitenden Betreuung und zielgerichteten Feedbacks bedanken. Dank gilt meiner Familie, insbesondere meiner Schwester Silvana, die mir mit ihren ausgeprägten sprachlichen Fähigkeiten half, diese Arbeit abschließend abzurunden und von allfälligen Fehlern zu bereinigen. Ein herzliches Dankeschön auch an meine Frau Sarah, die mir immer mit aufmunternden Worten zur Seite stand und zu guter Letzt noch ein großes Dankeschön an unsere Zwillinge, die mir aufgrund ihrer bevorstehenden Ankunft zusätzliche Motivation schenkten und somit meinen Fokus für das Wesentliche aufrecht hielten.



# Inhalt

<b>1. Einleitung</b> .....	<b>9</b>
<b>2. Theoretischer Hintergrund</b> .....	<b>13</b>
2.1. Leistungsbeurteilung in der Volksschule .....	13
Funktionen der Leistungsbeurteilung .....	14
Bezugsnormen der Leistungsbeurteilung.....	15
2.2. Beurteilung durch Ziffernnoten .....	18
Soziale Bezugsnorm .....	20
Empirische Studien .....	21
2.3. Verbalbeurteilungen .....	25
Individuelle und sachliche Bezugsnormen .....	26
Empirische Studien .....	28
2.4. Beobachtungs- und Beurteilungsfehler bei der Leistungsbeurteilung.....	32
Der <i>Halo-Effekt</i> im schulischen Kontext.....	33
2.5. <i>Offenheit (Big-Five)</i> bei Lehrer*innen .....	36
<i>Offenheit</i> und Leistungsbeurteilung.....	37
2.6. Beurteilung durch Ziffernnoten versus Verbalbeurteilung.....	39
Zwischenfazit .....	39
Empirische Studien zur Präferenz der Leistungsbeurteilung: Beurteilung durch Ziffernnoten versus Verbalbeurteilungen .....	42
<b>3. Fragestellung und Hypothesen</b> .....	<b>46</b>

<b>4. Methode</b> .....	<b>49</b>
4.1. Studiendesign.....	49
4.2. Erhebungsinstrumente .....	50
Beurteilungspräferenz .....	50
<i>Offenheit (Big-Five, Satow, 2012)</i> .....	51
Beurteilungsfehler bei der Leistungsbeurteilung ( <i>Halo-Effekt</i> ): Benotung eines Aufsatzes .....	52
4.3. Stichprobe .....	55
<b>5. Ergebnisse</b> .....	<b>57</b>
5.1. Unterschiede in den Beurteilungspräferenzen je Schulklasse von Lehrer*innen und Laien .....	58
5.2. Der Zusammenhang zwischen <i>Offenheit</i> und Beurteilungspräferenz.....	62
5.3. Beurteilungsfehler bei der Benotung ( <i>Halo-Effekt</i> ) .....	63
<b>6. Diskussion</b> .....	<b>65</b>
6.1. Unterschiede in den Beurteilungspräferenzen je Schulklasse von Lehrer*innen und Laien .....	65
6.2. Der Zusammenhang zwischen <i>Offenheit</i> und Leistungsbeurteilungspräferenz...	68
6.3. Beurteilungsfehler bei der Benotung ( <i>Halo-Effekt</i> ) .....	69
<b>7. Limitationen</b> .....	<b>70</b>
<b>8. Fazit</b> .....	<b>72</b>
<b>9. Abbildungsverzeichnis</b> .....	<b>74</b>
<b>10. Tabellenverzeichnis</b> .....	<b>74</b>
<b>11. Literaturverzeichnis</b> .....	<b>75</b>

<b>12. Anhang</b> .....	<b>88</b>
Anhang A: Zusammenfassung .....	88
Anhang B: Abstract .....	89
Anhang C: Rekrutierungsmail Lehrpersonen (Volksschulen) .....	90
Anhang D: Rekrutierungsschreiben Laien (Online-Foren).....	91
Anhang E: Online-Fragebogen.....	92
Anhang F: Items-Offenheit ( <i>B5T, Satow, 2012</i> ).....	99
Anhang G: Begründungen für die Beurteilungspräferenzen (Lehrer*innen) .....	100
Anhang H: Begründungen für die Beurteilungspräferenzen (Laien) .....	104
Anhang I: Begründungen für die Nicht-Benotung von Peters Geschichte .....	107



# 1. Einleitung

Bereits in den 1970er-Jahren wurde an der Beurteilung durch Ziffernoten in Schulen Kritik geübt (Ingenkamp, 1971). Seither werden alternative Leistungsbeurteilungsmethoden in Schulen erprobt. In Österreich werden schon seit 1966 Schulversuche mit verbalen Leistungsbeurteilungen in Volksschulen durchgeführt (Rechnungshof, 2015).

Leistungsbeurteilungsmethoden im schulischen Kontext sind gut erforscht und es gibt Evidenz für die mangelhaften Gütekriterien sowie potenziell negativen Effekte auf die Lernmotivation von Schüler\*innen von Ziffernoten (Butler, 2005; Meece et al., 2006; Weber & Petermann, 2016). Verbalbeurteilungen werden hingegen, bei korrekter Anwendung, von Schüler\*innen als motivationsfördernd erlebt und können die Lernentwicklung von Schüler\*innen positiv beeinflussen (Black & Wiliam, 2010; Hattie et al., 2017; Hess, 2019; Rakoczy et al., 2019).

Trotz all dieser Erkenntnisse ist die bevorzugte Art der Leistungsbeurteilung immer noch ein polarisierendes Thema, mit Fürsprecher\*innen sowohl für die Beurteilung durch Ziffernoten sowie für die Verbalbeurteilung.

Entgegen der wissenschaftlichen Erkenntnisse wurde im Herbst 2019 österreichweit unter der derzeitigen Regierung die Wiedereinführung der verpflichtenden Leistungsbeurteilung mittels Ziffernoten ab dem Jahreszeugnis der zweiten Volksschulklassen veranlasst. Eine verbale Leistungsbeurteilung der Schüler\*innen ist weiterhin bis zum Ende des ersten Semesters der zweiten Schulklasse möglich. Dafür ist jedoch eine Übereinkunft der jeweiligen Eltern, innerhalb der ersten neun Wochen des Schuljahres, notwendig (SchuG §18a).

In der Bildungspsychologie nimmt der Primärbereich eine besonders wichtige Rolle ein. In den Volksschulen werden Erziehungs- und Bildungsauftrag das erste Mal von den Eltern weg hin zu einer Institution gegeben. Die Entwicklung von Selbstbewusstsein und Leistungsfähigkeit der Schüler\*innen wird durch die zuständigen Pädagog\*innen sowie dem vorhandenen Schulsystem maßgebend beeinflusst (Spinath, 2010). Die Wahrnehmung der eigenen Fähigkeiten der Schüler\*innen im Primärbereich wird deshalb auch zunehmend durch die Beurteilungen der Lehrer\*innen geprägt (Spinath & Spinath, 2005).

Beobachtungs- und Beurteilungsfehler können sowohl Leistungsbeurteilungen durch Ziffernnoten sowie Verbalbeurteilung beeinflussen (Häcker et al., 2013). Vor allem die Beurteilung einzelner Schüler\*innen basierend auf der sozial vergleichenden Leistungsbeurteilung durch Ziffernnoten ist anfällig für Beurteilungsfehler und wird von Effekten, wie bspw. dem *Halo-Effekt*, verzerrt (Cooper, 1981; Feeley, 2002). Dieser beschreibt die Tendenz, dass eigentlich unabhängige Eigenschaften einer Person in die jeweilige Beurteilung miteinfließen. Das führt in weiterer Folge dazu, dass kaum mehr sinnvolle Rückschlüsse auf die zugrundeliegende erbrachte Leistung und Fähigkeiten der Schüler\*innen möglich sind. (Häcker et al., 2013; Malouff & Thorsteinsson, 2016).

Am Zustandekommen der Schulleistung ist nicht nur die/der Schüler\*in, sondern auch die Lehrperson, der Unterricht, die Schule und die Gesellschaft beteiligt (BIFIE, 2019). Die Diversität der Schüler\*innen in den Volksschulen ist besonders hoch im Vergleich zu den nachfolgenden Schulformen. In den Volksschulen werden viele Förderbedarfe, wie Hochbegabungen und Lernschwächen, festgestellt (BIFIE, 2019; Spiel & Wagner, 2003). Die ohnehin heterogene Zusammensetzung der 6- bis 10-Jährigen in den Volksschulen, vor allem in den urbanen Regionen, wird durch internationale Zuwanderung und Fluchtbewegungen vielfältiger (BIFIE, 2019). Gleichzeitig werden Anforderungen zur Internationalisierung, neue Lernformen und vermehrte Motivations- und Konzentrationsprobleme der Schüler\*innen zu wachsenden Belastungen der Lehrpersonen (Spiel & Wagner, 2003).

Lehrer\*innen werden mit diesen vielseitigen Problemen täglich konfrontiert und müssen daher immer wieder angepasst auf neue und individuelle Situationen im Unterricht reagieren. Personen mit einer hohen Ausprägung in *Offenheit* (*Big-Five*, Satow, 2012) gehen gerne auf neue Herausforderungen zu und sind im pädagogischen Kontext besser in der Lage, ihre Schüler\*innen aus verschiedenen Blickwinkeln zu betrachten und zu beurteilen (Hartinger et al., 2005; Faust und Foerster, 2008; Tan et al., 2019; Tatalović Vorkapić, 2012). Die größte Differenz im Vergleich der *Big-Five* Persönlichkeitseigenschaften zwischen Lehrpersonen und Laien (keinen Lehrberuf ausübende Personen) macht die *Offenheit* aus (Eulenberger, 2015). Das könnte für

einen Zusammenhang einer ausgeprägten *Offenheit* und der Präferenz für individuelle Verbalbeurteilungen gegenüber den sozial vergleichenden Ziffernnoten sprechen.

In sämtlichen Schulformen nach der Grundschule gibt es kaum Aufwärtsmobilität. Dies bedeutet, dass ein Aufstieg in eine höhere Schule wesentlich unwahrscheinlicher ist als ein Abstieg (Spiel et al., 2010). Daraus lässt sich ableiten, dass die Hauptaufgabe einer Grundschule und somit der Lehrer\*innen darin besteht, den Schüler\*innen ein positives Selbstbild zu vermitteln und sie mit einer differenzierten Selbsteinschätzung auszustatten. Um die Kinder auf die kommenden Ansprüche weiterer Ausbildungen sowie die Arbeitswelt vorzubereiten und positive kumulative (Lern-)Effekte zu erlangen (BIFIE 2019; Havighurst, 1972) sowie um sie bei der Entwicklung selbstständigen Denkens zu unterstützen, ist die Vergabe von Ziffernnoten als Leistungsbeurteilung nicht ausreichend (Deutsche Kultusministerkonferenz, 2007).

Die Ergebnisse von empirischen Studien sprechen vorwiegend für die positiven Effekte von Verbalbeurteilungen und der Reformation alter Schulsysteme mit Beurteilung durch Notengebung. Es zeichnet sich vor allem hinsichtlich der Grundschule eine Präferenz für Verbalbeurteilungen seitens Eltern, Lehrer\*innen und Schüler\*innen ab (Lütgert et al., 2001; Valtin et al., 2002).

Die vorliegende Arbeit beschäftigt sich mit Leistungsbeurteilungen in der Volksschule. In Kapitel 2.1 wird näher auf die Funktionen der Leistungsbeurteilung und die der Bewertungen zugrundeliegenden Bezugsnormen eingegangen. Die Beurteilung durch Ziffernnoten und die Verbalbeurteilung werden in den Kapiteln 2.2 und 2.3 mit ihren Charakteristika sowie Vor- und Nachteilen dargestellt. In Kapitel 2.4 wird näher auf Beurteilungs- und Beobachtungsfehler, insbesondere auf den *Halo-Effekt*, eingegangen. In Kapitel 2.5 wird der Zusammenhang der Persönlichkeitseigenschaft *Offenheit (Big-Five)* mit dem Lehrberuf und der präferierten Leistungsbeurteilung betrachtet. Den theoretischen Teil abschließend erfolgt in Kapitel 2.6 ein Zwischenfazit der vorgestellten Literatur betreffend die Zusammenschau der beiden Leistungsbeurteilungsvarianten und eine Vorstellung zweier groß angelegter empirischer Studien, die sich mit Vergleichen der Beurteilung durch Ziffernnoten und Verbalbeurteilung beschäftigen.

Darauffolgend werden in Kapitel 3 die Fragestellung sowie Hypothesen und in Kapitel 4 die Methode der Studie dargestellt. Anschließend werden die Ergebnisse entlang der formulierten Hypothesen präsentiert sowie weitere Befunde der Untersuchung berichtet (Kapitel 5). Abschließend folgen in Kapitel 6 die Diskussion zu den Ergebnissen, die Limitationen der Untersuchung (Kapitel 7) und das Fazit (Kapitel 8).

## 2. Theoretischer Hintergrund

Am Beginn dieses Kapitels steht die Erklärung der notwendigen Hintergründe und allfälliger Begrifflichkeiten für die Untersuchung. Anschließend wird näher auf die schulischen Leistungsbeurteilungsmethoden (Beurteilung mit Ziffernnoten und Verbalbeurteilung), Beobachtungs- und Beurteilungsfehler bei der Leistungsbeurteilung und *Offenheit* eingegangen. Abschließend werden zwei empirische Studien zur Gegenüberstellung der Leistungsbeurteilung mit Ziffernnoten versus Verbalbeurteilungen vorgestellt und damit zu der Fragestellung und den Hypothesen übergeleitet, bevor auf den empirischen Untersuchungsteil eingegangen wird.

### 2.1. Leistungsbeurteilung in der Volksschule

Schulische Lernerfahrungen von Kindern sind grundlegende Einflussfaktoren für positive Selbstkonzepte, Interessen und der Freude am Lernen neuer Dinge (Schreiner et al., 2018). Laut Pekrun (1997) weisen Persönlichkeitsmerkmale bereits nach der Grundschule ein hohes Maß an Stabilität auf. Es ist also notwendig, bereits in der Volksschule, eine bestmögliche, differenzierte Unterstützung einzelner Schüler\*innen zu gewährleisten. Im Laufe der Schulzeit in der Primarstufe zeigen die Schulleistungen ein großes Veränderungspotential (Weinert & Helmke, 1997). Der Einfluss motivationaler Persönlichkeitsmerkmale für die allgemeine kognitive Entwicklung und schulischen Leistungen von Schüler\*innen ist bereits bekannt und von beträchtlicher Bedeutung (Heller, 1997).

## ***Funktionen der Leistungsbeurteilung***

Beurteilungen und Bewertungen von Leistungen der Schüler\*innen gehören zu den alltäglichen Anforderungen an Lehrer\*innen und bilden die Grundlage zur Optimierung von Lehr-Lern-Prozessen (Weinert, 2001). Den Rahmen für diese Bewertungen bietet die Leistungsbeurteilungsverordnung (LBVO, 1974), sie definiert auch die Funktionen der Leistungsbeurteilung. Es wird zwischen pädagogischen, gesellschaftlichen und personenbezogenen Funktionen unterschieden.

Pädagogische Funktionen: Der Inhalt zur Vorbereitung auf Prüfungen soll Signalfunktionen für die Schüler\*innen darbieten und somit den Lehr-Lern-Prozess steuern. Wenn bspw. Leistungsanforderungen für die Schüler\*innen nachvollziehbar sind und auf einem hohen taxonomischen Niveau überprüft werden, wird den Schüler\*innen nahegelegt, dass es um elaboriertes Wissen und nicht um bloßes Abrufen von Informationen geht. Die Feedbackfunktion dient der Verbesserung von Lernwegen und zur Fehleranalyse. Feedback erfolgt als Reaktion auf erbrachte Leistungen, um mittels diagnostischer Informationsweitergabe, Lernprozesse zu optimieren (Bürgermeister & Saalbach, 2018). Adäquates Feedback gibt Informationen über den aktuellen Lernstand sowie das zu erreichende Lernziel und zeigt brauchbare, individuelle Handlungsmöglichkeiten auf, um das Lernziel zu erreichen (Winter, 2015).

Die Motivierungs- und Disziplinierungsfunktion von Ziffernnoten ist umstritten, weil die Gefahr für negative Auswirkungen auf die Lernmotivation und das Selbstkonzept der Schüler\*innen besteht.

Gesellschaftliche Funktionen: Schüler\*innen werden innerhalb der Schule mittels der Leistungsbeurteilung klassifiziert, um für eine Bildungslaufbahn, bspw. eine höhere Schule, zugewiesen werden zu können (Allokationsfunktion). Die Selektion erfolgt schlussendlich mithilfe von Schulnoten. Die Berichtsfunktion der Leistungsbeurteilung dient der nachvollziehbaren Informationsweitergabe über eine erfolgreiche Absolvierung schulischer Anforderungen für die Schüler\*innen selbst, den Eltern und Abnehmer\*innen außerhalb der jeweiligen Schule.

Personenbezogene Funktionen: Die Leistungsbeurteilung beeinflusst die Einstellung zur Schule und das Selbstkonzept, was wiederum die weiteren Bildungsaspirationen beeinflusst. Durch die Leistungsbeurteilung können Schul- und Prüfungsangst sowie die (lebenslange) Lernmotivation maßgebend beeinflusst werden (Eder et al., 2009; LBVO, 1974).

Im folgenden Kapitel wird näher auf die den schulischen Leistungsbeurteilungen zugrundeliegenden Bezugsnormen eingegangen.

### ***Bezugsnormen der Leistungsbeurteilung***

Leistungsmessungen und -bewertungen werden in vielfältigen Formen in Psychologie und Pädagogik angewendet. Ein (psychologischer) Test dient bspw. dazu, Ausprägung einer Fähigkeit, Eigenschaft oder eines Kenntnisstandes zu erörtern und/oder zukünftiges Verhalten vorherzusagen. Ob ein Test eine ausreichend hohe Qualität aufweist, dem liegt eine komplexe Entscheidung zugrunde. Ein wichtiger Indikator für einen qualitativ hochwertigen Test ist das ausreichende Vorhandensein der qualitätssichernden Gütekriterien: Objektivität, Reliabilität, Validität und Normierung (Häcker et al., 2013; Häcker et al., 1998).

Es gibt drei verschiedene Bezugsnormen zur Beurteilung von Schulleistungen, wenn das Anwenden von standardisierten Tests nicht möglich ist bzw. keine standardisierten Tests zur Verfügung stehen, wie es im schulischen Alltag oft der Fall ist (Häcker et al., 2013; Rheinberg, 2006):

Individuelle Bezugsnorm: Individueller Lernzuwachs oder -rückschritt rückt hier in den Fokus der Bewertung. Dies kann für schwächere Schüler\*innen, die Lernerfolge erzielen, durchaus einen motivierenden Effekt aufweisen und zu einer positiveren Selbsteinschätzung führen. Misserfolgs- und Prüfungsängste werden verringert und größere Lernfreude – eine zeitlich relativ stabile emotionale Besetzung des schulischen bzw. akademischen Lernens und fachbezogener Inhalte (Helmke, 1993) – sowie Mitarbeit können verstärkt werden (Rheinberg, 1980). Nachteilig ist hier die fehlende

Selektionsmöglichkeit, da keine Vergleiche mit anderen Schüler\*innen vollzogen werden.

Soziale Bezugsnorm: Individualleistungen werden in Abhängigkeit zur durchschnittlichen Klassenleistungsfähigkeit beurteilt. Diese relativen Individualleistungen ermöglichen es, Rückschlüsse auf Förderbedarfe der besseren sowie der schlechteren Schüler\*innen zu ziehen. Nachteile dieser Norm bestehen, weil Kinder mit gleichen erzielten Leistungen je nach Referenzklasse einer unterschiedlichen Beurteilung unterzogen werden könnten. Es kann ein zu langsames, aber auch rückschrittliches Lernen stattfinden, weil der absolute Lernzuwachs nicht als Vergleichskriterium herangezogen wird.

Sachliche Bezugsnorm: Ein vorgegebener inhaltlicher Leistungsstandard dient als Bewertungskriterium der einzelnen erbrachten Leistungen. Diese haben den Vorteil, leicht aufzuzeigen, ob bestimmte Mindestanforderungen erreicht wurden. Mittlerweile werden Lehrpläne mithilfe von Bildungsstandards erstellt, um klarzustellen, welche Kompetenzen in welchen Klassenstufen und Unterrichtsfächern erfüllt werden sollten. Diese Methode ist wiederum nicht frei von der sozialen Bezugsnorm, da diese für die Normierung von Tests notwendig ist (Häcker et al., 2013; Rheinberg, 2006; Rheinberg et al., 2001). Unter „Bezugsnormorientierung“ versteht man den tendenziell bevorzugten Vergleichsstandard von Lehrpersonen bei Leistungsbewertungen von Schüler\*innen (Rheinberg, 1980). Empirische Befunde sprechen für Auswirkungen der verschiedenen Bezugsnormorientierungen auf die Lernleistung und motivational-affektive Aspekte des Lernens (Rheinberg, 2008). Laut der Untersuchung von Fischbach et al. (2016) unterscheiden sich Lehrpersonen in der zur Beurteilung von Schüler\*innenleistungen herangezogenen Bezugsnormen. Eine individuelle Bezugsnormorientierung wirkt sich positiv auf Lernmotivation sowie Leistung von Schüler\*innen aus: Geringere Ängstlichkeit in Prüfungssituationen und vor Misserfolgen, realistischere Zielsetzungen und Förderung von Leistungsselbstkonzepten sind der Fall (Köller, 2004; Lüdtke et al., 2005; Mischo & Rheinberg, 1995). Eine soziale Bezugsnormorientierung bei der Leistungsbeurteilung wirkt sich, wegen der sozialen Vergleiche, potenziell negativ auf die Lern- und Schulleistungen aus (Fischbach et al., 2016).

Die positiven und negativen Effekte der Bezugsnormorientierungen sind bei leistungsschwächeren Schüler\*innen stärker ausgeprägt als bei leistungsstärkeren Schüler\*innen (Fischbach et al., 2016; Jürgens & Sacher, 2008).

Unabhängig von der Bezugsnormorientierung der jeweiligen Lehrperson entwickeln Schüler\*innen ihre leistungsbezogenen Selbstkonzepte auch im sozialen Vergleich mit ihrer Peergruppe (Möller et al., 2009). Eine ausgeprägte individuelle Bezugsnorm der Lehrperson kompensiert also nicht zwingend die (negativen) Effekte auf das Selbstkonzept durch soziale Vergleiche zwischen den Kindern (Lüdtke et al., 2005).

Bezugsnormorientierung beeinflusst zwar das Lernverhalten, Schüler\*innen stellen aber trotzdem unabhängig von der durch die Lehrperson verwendeten Bezugsnorm interindividuelle Vergleiche mit ihren Mitschüler\*innen an (Fischbach et al., 2016).

Die Art der angewendeten Leistungsbeurteilung beeinflusst die Leistungswahrnehmung und somit die positiven und negativen Einflüsse der unterschiedlichen Bezugsnormen. So wird durch die Beurteilung durch Ziffernnoten eine soziale Bezugsnorm forciert (Marsh et al., 2014), wohingegen durch eine verbale Beurteilung schulischer Leistungen eine individuelle und sachliche Sichtweise bevorzugt wird (Black & Wiliam, 2010).

In den folgenden Kapiteln wird näher auf die Beurteilungsmethoden „Beurteilung durch Ziffernnoten“ und „Verbalbeurteilungen“ eingegangen und deren zugrundeliegenden Charakteristika, Bezugsnormen sowie empirischen Untersuchungen betreffend die Vor- und Nachteile der Methoden dargestellt.

## 2.2. Beurteilung durch Ziffernnoten

Nach der österreichischen Leistungsbeurteilungsverordnung (LBVO, 1974) erfolgt die Notenvergabe in zwei Phasen, „Leistungsfeststellung“ gefolgt von der „Leistungsbeurteilung“. Erstens werden Schüler\*innenleistungen durch die Messung von Lernergebnissen ermittelt. Dazu werden Mitarbeit sowie mündliche, schriftliche, praktische oder graphische Leistungsnachweise herangezogen. Zweitens werden die erbrachten Leistungen mithilfe eines Beurteilungsmaßstabes in Ziffernnoten zusammengefasst (LBVO, 1974; Neuweg, 2014). Neben den historisch relevantesten Qualifikationen – Lesen, Schreiben und Rechnen – werden Wert-, Arbeitshaltung und Sozialkompetenz im Umgang mit Gleichaltrigen durch die Lehrpersonen bewertet. Jede Lehrkraft entscheidet selbständig über die Einflussnahme dieser Kategorien auf die jeweilige Ziffernnote (Spiel & Wagner, 2003).

In Volksschulen erfolgt seit der Bildungsreform 2018 die Leistungsbeurteilung durch Ziffernnoten ab dem Jahreszeugnis der zweiten Klasse. In Tabelle 1 wird das Beurteilungsspektrum dargestellt (LBVO §14; SchUG §18 Abs. 2 und 3).

Tabelle 1: §14 der LBVO:

	<b>Reproduktive Leistung</b>	<b>Produktive Leistung</b>	
	Erfassung und Anwendung des Lehrstoffes; Durchführung der Aufgaben	Eigenständigkeit	Selbstständige Anwendung des Wissens und Könnens auf neuartige Aufgaben
<b>Sehr gut</b>	In weit über das Wesentliche hinausgehendem Ausmaß	Deutlich (wo dies möglich ist)	Muss vorliegen (wo dies möglich ist)
<b>Gut</b>	In über das Wesentliche hinausgehendem Ausmaß	Merklich (wo dies möglich ist)	Bei entsprechender Anleitung (wo dies möglich ist)
<b>Befriedigend</b>	In den wesentlichen Bereichen zur Gänze erfüllt	Mängel in der Durchführung der Aufgaben werden durch merkliche Ansätze ausgeglichen	Bei entsprechender Anleitung (wo dies möglich ist)
<b>Genügend</b>	In den wesentlichen Bereichen überwiegend		
<b>Nicht Genügend</b>	Nicht einmal in den wesentlichen Bereichen überwiegend		

Tabelle 1: Notenskala: 1 = „Sehr gut“, 2 = „Gut“, 3 = „Befriedigend“, 4 = „Genügend“, 5 = „Nicht genügend“ laut LBVO §14 (Neuweg, 2014: S. 74)

## **Soziale Bezugsnorm**

Die Beurteilung durch Ziffernnoten erfüllt pädagogische und gesellschaftliche Funktionen (Lintorf, 2011; Tent & Birkel, 2010). Als ein pädagogischer Aspekt gilt die Rückmeldefunktion an die Schüler\*innen, um das Lernverhalten anzupassen (Tillmann & Vollstädt, 2000). Die Selektions- und Allokationsfunktionen zählen zu den gesellschaftlichen Aspekten (Fend, 2008). Aufgrund der sozialen Bezugsnorm, bzw. klasseninternen Vergleichen der Schüler\*innen untereinander, kann es bei Schüler\*innen mit schlechteren Schulnoten zu einem geringeren Selbstwert führen (Marsh, 1987; Marsh et al., 2014). Das wiederum kann zu verminderter Lernfreude und erhöhter Leistungsangst führen (Zeinz & Köller, 2006).

Ziffernnoten weisen in sämtlichen diagnostischen Gütekriterien Mängel auf. Objektivität wird aufgrund der klasseninternen, sich auf die soziale Bezugsnorm beziehenden Vergleiche nicht erreicht. Außerdem liefern Noten nur wenige Informationen über Kompetenzen, Teilbereichswissen und Lernstrategien der einzelnen Schüler\*innen, sondern vergleichen sie nur in Relation mit ihrem Klassenverband. Deshalb ist es nicht möglich, das Zustandekommen einer Ziffernnote zu bestimmen (Valtin & Rosenfeld, 1997).

Die aggregierten Jahresnoten entstehen wiederum durch auf verschiedenen Grundlagen zusammengefassten Prüfungsnoten. Problematisch sind zusätzlich auch die Tatsachen, dass nicht jede Einzelprüfung eine vergleichbare Aufgabenvielfalt bietet und mit ordinalskalierten Noten keine arithmetischen Mittelwerte berechnet werden sollten (Harwell & Gatti, 2001). Dadurch wird die Nachvollziehbarkeit von Jahresnoten beinahe unmöglich (BIFIE, 2019). So kommt es, die Leistungsspannen betreffend, zwischen verschiedenen Notenstufen zu großen Überschneidungen und innerhalb von Notenstufen zu hohen Streuungen (Paasch et al., 2019).

Laut Valtin und Rosenfeld (1997) wurden Ziffernnoten und -zeugnisse aufgrund der praktischen Allokations- und Kontrollfunktionen und nicht wegen pädagogischer Zweckmäßigkeit eingeführt. Das ist im Hinblick auf die Notenverteilung in den vierten Klassen der Primarstufe erkennbar. Die mit großem Abstand am häufigsten vergebenen

Noten in den Fächern Mathematik und Deutsch sind „Sehr gut“ und „Gut“ (Paasch et al., 2019). Diese Notenverteilung ist einzigartig in Grundschulen und könnte daran liegen, dass die Allokationsfunktion der Ziffernnoten einen erheblichen Einfluss auf die Leistungsbeurteilungen durch die Lehrpersonen nimmt. Eine Beurteilung mit der Note „Befriedigend“ gilt als Ausschlusskriterium für den Aufstieg in die allgemeinbildenden höheren Schulen (AHS), so kann die Vergabe schlechterer Beurteilungen („Genügend“ und „Nicht genügend“) umgangen werden, um die Schüler\*innen nicht zusätzlich zu stigmatisieren. Diese Beurteilungstendenzen sind sowohl in Österreich als auch in Deutschland zu finden und sind auf die soziale Bezugsnorm der Beurteilung durch Ziffernnoten zurückzuführen (Schmid et al., 2016).

### ***Empirische Studien***

In den statistischen Berechnungen der BIFIE (2019) wird nur ein geringer Zusammenhang von Leseverständnis und Deutschnoten in den vierten Klassen der Sekundarschulen berichtet. Dasselbe Phänomen tritt bei Mathematikleistungen und -noten in den achten Schulstufen der Klassen der AHS auf. Eine transparente und kriteriale Benotung könnte zu passenderen Rückmeldungen des jeweiligen Leistungsstandes führen und die Beurteilungsgerechtigkeit verbessern. Diese Umstellung könnte aber eine Verschlechterung des Notenspiegels nach sich ziehen und somit eine negative Auswirkung auf die Übergangsquoten mit sich bringen. Viele Schüler\*innen hätten somit geringere Bildungschancen, was mit einer niedrigeren Lernmotivation und einem geringeren Selbstkonzept einhergehen würde. Gesellschaftlich betrachtet würde das Bildungspotenzial nicht ausgeschöpft werden, weil weniger Schüler\*innen höhere Schulen besuchen könnten (Paasch et al., 2019).

Aufgrund der wechselseitigen Beziehung von Freude am Unterricht und Fähigkeitsselbstkonzept der Schüler\*innen, sollte die Aufrechterhaltung der Lernfreude ein primäres Unterrichtsziel sein (Schreiner et al., 2018). Bereits erlebte Situationen im Kontext des schulischen Wissenserwerbes sowie Wissensüberprüfungen durch Lehrkräfte beeinflussen zukünftige Leistungsemotionen. So bestimmen positive Emotionen unter anderem darüber, ob elaborierende Lernstrategien oder hohe

Konzentration stattfinden (Hascher & Edlinger, 2009). Wiederum negative Emotionen wie Prüfungsangst oder Ärger können leistungsbeeinträchtigend auf das Lösen schwieriger Aufgaben wirken. So kommt es bei hochängstlichen Schüler\*innen trotz vorhandenem Vorwissen und ausreichenden kognitiven Fähigkeiten, im Gegensatz zu niedrigängstlichen Schüler\*innen, zu schlechteren Leistungen in nahezu sämtlichen Schulfächern. Ängstliche Kinder konzentrieren sich in Prüfungssituationen häufiger auf sich selbst und ihre Mängel, anstatt ihren Fokus auf das zu lösende Problem zu lenken (Götz et al., 2013). Laut Eder (2009) hätten mehr als 20 % der Schüler\*innen oft Prüfungsangst. Die empfundene Angst steigt am Ende der Primarstufe deutlich an und bleibt dann in der weiterführenden Sekundarstufe in allen Schultypen annähernd gleich hoch ausgebildet. Yerkes und Dodson (1908) haben bereits vor über 100 Jahren ein Modell kreiert, welches den Zusammenhang von Leistungspotenzial und gegebenem Erregungsniveau darstellt. Es wird davon ausgegangen, dass eine hohe kognitive Aktivierung bei schwierigen Aufgaben nur bei einem geringen Erregungsniveau möglich ist und somit eine valide Darstellung der Leistung einzelner Schüler\*innen mit hoher Ängstlichkeit bei schwierigen Aufgaben nicht möglich ist. Durch häufige schulische Misserfolge kann ein negativ verzerrtes Bild der eigenen Fähigkeiten und in weiterer Folge eine Vermeidungshaltung entstehen. Über die Schulzeit kann dies zu kumulierten Fähigkeitsdefiziten führen (Helmke, 1993). Lohbeck et al. (2016) untersuchten 178 Grundschulkindern der vierten Schulklasse hinsichtlich der Zusammenhänge ihrer Lernfreude, Schulleistungen, schulischen Selbstkonzepte und Konzentration. Es zeigten sich direkte Korrelationen zwischen schulischem Selbstkonzept und Schulleistung; Selbstkonzept und Lernfreude; Selbstkonzept und Konzentration; Lernfreude und Konzentration. Schüler\*innen vergleichen ihre Leistungen internal (verschiedene Fächer) sowie external (mit ihren Mitschüler\*innen) (Möller und Marsh, 2013). Durch die externen Vergleiche entstehende Kontrasteffekte erklären, warum es bei relativ besseren Leistungen zu einem positiveren Selbstbild und bei relativ schlechteren Leistungen zu einem negativeren Selbstbild kommen kann (Lohbeck et al., 2016).

Um den schulischen Ansprüchen gerecht zu werden, ist die Erhaltung und Förderung der Leistungsbereitschaft der Schüler\*innen durch konstruktive und schlüssige Leistungsrückmeldungen von großer Wichtigkeit. Dadurch werden den Kindern Interpretationsmöglichkeiten für Erfolge sowie Misserfolge dargeboten, die ihre zukünftigen schulrelevanten Persönlichkeitsmerkmale in ihrer Entwicklung prägen. Dabei spielen Resultat, Prozess und die zugrundeliegende Bezugsnorm zentrale Rollen (Oerter, 1998).

Ziffernnoten beurteilen hauptsächlich die erzielten Ergebnisse anstatt der zugrundeliegenden Lösungsstrategien und verstärken somit noch zusätzlich die sozialen Vergleiche der Schüler\*innen untereinander (Schmude, 2005). Sie legen nahe, dass Erfolg mit fehlerfreiem Können gleichzusetzen ist und sich nur mit Noten messen lässt. Somit gerät der Lernprozess in den Hintergrund und verliert an Relevanz. Das Projekt NOVUS (Noten oder Verbalbeurteilung: Unterrichtsorganisation und Sanktionsverhalten von Lehrkräften in Ost- und Westberliner Grundschulen; Wagener, 2003) bestätigte negative Auswirkungen von Ziffernnotenbeurteilungen. Schwächere Schüler\*innen entwickelten die Einstellung, die schulischen Anforderungen nicht erfüllen zu können und bekamen, durch die Rangreihung der Ziffernnoten, ihren Platz mittels der leistungsbezogenen Platzierung innerhalb des Klassenverbandes zugewiesen. Die Münchner LOGIK-Studie (Longitudinalstudie zur Genese individueller Kompetenzen; Helmke & Weinert, 1999; Schneider, 2008) untersuchte Zusammenhänge zwischen der Entwicklung von kognitiven, sozialen sowie motorischen Kompetenzen und der Entwicklung von schulischen Fähigkeiten. Sie wurde im Zeitraum von 1984 bis 2005 bei etwa 200 vierjährigen Kindern durchgeführt. Das leistungsbezogene Selbstkonzept der Kinder mit den Schwerpunkten Deutsch und Mathematik wurde fortlaufend im Kindergarten sowie in der Grundschule mithilfe von Interviews und Fragebögen erfasst. Wie aus der Literatur bereits bekannt, stellt sich im Verlauf der Grundschule eine realistischere und stabilere - meist negativere - Selbsteinschätzung der eigenen Fähigkeiten ein. Dies könnte u.a. an der in diesem Zeitraum stattfindenden Einführung der Notenbeurteilungen liegen, welche sich zusätzlich negativ auf das Selbstkonzept der Schüler\*innen auswirken könnte (Schneider & Bullock, 2008; Spiel & Wagner, 2003).

Laut Butler (2005) beeinflusst die Art der Leistungszielorientierung die Interpretation von Feedback bzw. erhaltenen Schulnoten. Lernzielorientierte Schüler\*innen sind erpicht darauf, ihre Fähigkeiten zu verbessern und suchen - entweder um ihre Leistung zu demonstrieren oder sie zu verbergen - den Vergleich mit ihren Mitschüler\*innen. Im Zusammenhang mit der erlangten Ziffernnote, nehmen die Kinder Fehler als selbstwertbedrohliche Mängel ihrer Fähigkeiten wahr. Weil Schulnoten von Schüler\*innen oft als Hinweis für stabile Fähigkeiten gesehen werden, können sich schlechte Schulnoten auch negativ auf den Selbstwert der jeweiligen Schüler\*innen auswirken. Somit forciert das inhaltsleere Feedback von Schulnoten das Konkurrenzdenken der Schüler\*innen untereinander, anstatt sie zu ermutigen, individuellen Leistungszielen zu folgen (Meece et al., 2006).

Laut der Zusammenhangsanalyse von Schulangst, -unlust, Anstrengungsvermeidung und Noten in den Schulfächern Mathematik und Deutsch von Weber und Petermann (2016) wurden geringe bis mittlere Korrelationen gefunden. Die am höchsten ausgeprägten Interkorrelationen wurden zwischen Prüfungsangst sowie Anstrengungsvermeidung und Schulnoten gefunden. Darüberhinausgehend erlauben Ziffernnoten keine ausreichend detaillierten Rückmeldungen an die Eltern der jeweiligen Kinder. Vor allem Eltern mit niedrigem Bildungsstand sind nicht in der Lage, adäquate Konsequenzen aus den Rückmeldungen zu ziehen, um die Bildungschancen ihrer Kinder zu verbessern, was wiederum zu einem niedrigeren Selbstbewusstsein der Schüler\*innen führen kann (Filippin & Paccagnella, 2012).

## 2.3. Verbalbeurteilungen

In der LBVO §23a und im SchuG §18a wird beschrieben, was bei verbalen Leistungsbeurteilungen in Volksschulen bis einschließlich des ersten Semesters der zweiten Klasse berücksichtigt werden soll. Schüler\*innen sollen hinsichtlich „der Erfassung und der Anwendung des Lehrstoffes, der Durchführung der Aufgaben, der Eigenständigkeit und der festgestellten Lernfortschritte, Leistungsstärken, Begabungen und allfälligen Mängel, gemessen an den Lernzielen“ bewertet werden (LBVO §23a). Sinngemäß sind die eben genannten, zur Beurteilung heranzuziehenden, Aspekte auch im SchuG §18a zu finden. Naturgemäß gibt es bei den Verbalbeurteilungen, im Gegensatz zu den Ziffernnoten, individuellere Beurteilungsmöglichkeiten. So werden für die Lernstandsentwicklung Lernjournale und -portfolios zur Dokumentation herangezogen. Sie dienen der Nachvollziehbarkeit des prozeduralen Verlaufes und nicht nur einer einfachen summativen Beurteilung (Stern, 2010).

Aus bestehender Literatur lassen sich praxisrelevante Prinzipien zur Anwendung von Verbalbeurteilungen ableiten. Bspw. muss nach Hattie et al. (2017) „Lernen sichtbar gemacht“ werden, damit es für Schüler\*innen einen lernförderlichen Nutzen hat. Die zur Leistungsbeurteilung herangezogenen Informationen sollen nicht aus zufälligen Beobachtungen bestehen, sondern bereits vorher kategorisch festgelegt werden. Der Output der Schüler\*innen soll nicht nur hinsichtlich der Lernergebnisse, sondern auch betreffend die Prozessqualität des Lernfortschritts bewertet werden. Weiters kann eine adäquate Verbalbeurteilung eine „dialogisch-reflexive“ Feedbackbasis zwischen Lehrpersonen, Schüler\*innen und Peers bieten (Winter, 2004). Die Leistungsbewertung wird so zu einem integrierten und aktiven Prozess im Unterricht und es findet im Gegensatz zur Beurteilung durch Ziffernnoten keine Abgrenzung zur Prüfungssituation statt. Dieser Lernprozess kann in Lerntagebüchern festgehalten werden und dient somit der laufenden Reflexion der individuellen Fähigkeiten der Schüler\*innen (Vierlinger, 2012). Die verbalen Leistungsbeurteilungen sollen mit qualitativen Inhalten (bspw. „Ziel

erreicht“) oder einer Erläuterung des gegebenen Lernstands erfolgen (Winter, 2015; Winter, 2004).

Das aufgrund der Leistungsbeurteilung verwendete individuelle Feedback hat an den Wissensstand und das Können der Schüler\*innen angepasst zu werden, weil ein und dieselbe Rückmeldung für eine/n Schüler\*in förderlich, für eine/n andere/n Schüler\*in wiederum unbrauchbar oder sogar nachteilhaft sein kann (William, 2013). In Tabelle 2 ist eine Veranschaulichung förderlicher Gestaltung von Verbalbeurteilungen, mit den hervorzuhebenden Kerninhalten „Lernziel“, „Lernstand“ und „Lernweg“ dargestellt.

Tabelle 2: Fünf Schlüsselstrategien der formativen Leistungsbeurteilung:

	<b>Lernziel</b> (Wohin die/der Lernende geht)	<b>Lernstand</b> (Wo die/der Lernende steht)	<b>Lernweg</b> (Wie sie/er dorthin kommt)
Lehrperson	Lernziele und Erfolgskriterien mit den Lernenden klären	Effektive Diskussionen, Aktivitäten und Aufgaben arrangieren, die zu beobachtbaren Lernergebnisse führen	Feedback geben, dass das Lernen voranbringt
Peer		Schüler*innen ermöglichen, einander als Ressourcen zu nutzen	
Lernende		Schüler*innen als Verantwortliche ihres eigenen Lernens anerkennen	

Tabelle 2: Fünf Schlüsselstrategien der formativen Leistungsbeurteilung (William, 2014, S.8).

### **Individuelle und sachliche Bezugsnormen**

Im Gegensatz zu Ziffernnoten erlauben verbale Leistungsbeurteilungen ein grundsätzlich differenzierteres Feedback von Schulleistungen. Diese orientiert sich (intentionsgemäß) an individuellen und sachlichen Bezugsnormen. Diese Art der Leistungsbewertung erlaubt es, schon kleine Fortschritte im Lernprozess der Schüler\*innen zu kommunizieren (Rheinberg, 2006). Somit werden mithilfe der individuellen und sachlichen Bezugsnorm bei jeder Beurteilung sowohl das Ausmaß der kriterialen Zielerreichung und die intraindividuellen Lernfortschritte rückgemeldet (Jürgens & Sacher, 2008).

Der positive Einfluss der individuellen Bezugsnorm auf das aktuelle Leistungsverhalten und die Entwicklung der Leistungsmotivation im Sinne einer stärkeren Erfolgsorientierung ist empirisch belegt. So stellten Harks et al. (2014) in ihrer Untersuchung Vorteile im Zusammenhang mit individueller und sachlicher Bezugsnorm im Gegensatz zur sozialen Bezugsnorm fest. Schüler\*innen, die Leistungsbewertungen auf Basis der individuellen Bezugsnorm (also die intraindividuelle Entwicklung betreffend) erhielten, wiesen im Durchschnitt eine begünstigtere Motivationsstruktur auf und hatten seltener Misserfolgsangst als Kinder, die aufgrund der sozialen Bezugsnorm, mittels Ziffernnoten, beurteilt wurden. Die Kinder attribuierten Misserfolge seltener auf zeitlich überdauernde individuelle Persönlichkeitseigenschaften und hatten, aufgrund des Leistungsziel- und prozessorientierten Feedbacks, eine positivere Einstellung zu schulischem Lernen und ihre Fähigkeitsselbstschätzung verbesserte sich, was wiederum motivationssteigernde Effekte hatte. Durch verbale Beurteilungen der erbrachten Leistungen von Schüler\*innen sind detailliertere Evaluationen der Fähigkeiten, Stärken und Schwächen der Schüler\*innen möglich. Es wird der Lernprozess einzelner Lernender in den Blick genommen, dadurch gibt es nicht nur einen Vergleich der Leistungen innerhalb der Lerngruppe, sondern vorwiegend individuelle Zielsetzungen (Black & Wiliam, 2010). Mittels Klassenbeobachtungen konnten Noor et al. (2010) verbale Feedbacks von Grundschullehrer\*innen an Schüler\*innen untersuchen. Die verbalen Rückmeldungen durch die Lehrer\*innen hatten signifikanten Einfluss auf das Lernverhalten der Schüler\*innen. 63 % des gegebenen Feedbacks war evaluativ. Evaluative Rückmeldungen sind eine Art der Leistungsbeurteilung (Nunn, 2001) und werden in den meisten Fällen mit „sehr gut“, „gut“, „ja“, etc. ausgedrückt. Ein für die Schüler\*innen nachvollziehbares und nützliches Feedback sollte aber sowohl evaluativ als auch deskriptiv sein und somit dem/der Schüler\*in auch Möglichkeiten der individuellen Leistungsverbesserung aufzeigen (Wiliam, 2013). Weiters dienen motivationale Rückmeldungen, die erbrachte Leistung betreffend (bspw. die aufgewendete Mühe), einer verbesserten Lernmotivation (Dresel, 2010).

## ***Empirische Studien***

Smith und Gorard (2005) führten eine einjährige experimentelle Evaluation in einer Sekundarschule durch. 104 Schüler\*innen wurden in vier Gruppen von jeweils 26 Schüler\*innen eingeteilt: Eine Gruppe erhielt ein Jahr lang verbales Feedback ohne Noten, die anderen drei Gruppen erhielten, so wie es auf der Schule üblich war, in derselben Zeit Noten mit sehr wenig verbalem Feedback. Die Leistungsveränderungen der Schüler\*innen im Laufe des Studienzeitraumes wurden mittels kognitivem Fähigkeitstest und einem Englisch-Leistungstest aufgezeichnet. Die Studie kam zu dem Ergebnis, dass die Schüler\*innen in der Gruppe der Verbalbeurteilungen leicht schwächere Leistungen erbrachten als die Schüler\*innen der anderen drei Gruppen. Als Erklärung für dieses Ergebnis wurden mehrere Aspekte in Betracht gezogen. Die meisten Schüler\*innen der reinen Verbalbeurteilungsgruppe gaben in zwischenzeitlichen Befragungen an, wieder mit Noten beurteilt werden zu wollen. Die Begründungen dafür waren unterschiedlich, bspw. erwähnten sie, sich ausgegrenzt zu fühlen, weil sie Vergleichsmöglichkeiten mit den Mitschüler\*innen vermissten oder sie das erhaltene Feedback nicht praxisrelevant umsetzen konnten. Nur ein kleiner Anteil der Schüler\*innen in der Versuchsgruppe mit den Verbalbeurteilungen fand diese auch verständlich und hilfreich. Meistens zielten die Verbalbeurteilungen darauf ab, das Selbstbild und Selbstbewusstsein der Schüler\*innen zu stärken, nicht aber ihnen konkrete Hilfestellungen zur Verbesserung ihrer Leistung zu geben. Es wurde von den Schüler\*innen in den unterschiedlichen Versuchsgruppen der Konsens gefunden, dass parallel Noten- und Verbalbeurteilung die beste Variante sei. Dabei ist nach Black und Wiliam (2010) die dominante Präsenz der Ziffernnoten problematisch. Schüler\*innen würden zuallererst die Noten mit ihren Mitschüler\*innen vergleichen und die Verbalbeurteilungen würden wenig bis gar keine Beachtung finden. Somit ginge der eigentliche Nutzen der Verbalbeurteilungen verloren.

Laut Stobart und Gipps (1997) kann eine Leistungsbeurteilung nur dann förderlich sein, wenn sie am Lehr-Lern-Prozess ansetzt und die Lücke zwischen der aktuell erbrachten Leistung und der erwünschten Leistung beschreibt und sich mit Stärken und Schwächen der Schüler\*innen auseinandersetzt. Mit Ziffernnoten sei dies nicht möglich. Bei Verbalbeurteilung sei darauf zu achten, keine allgemeinen, sondern individuelle Aussagen zu tätigen. Eine angemessene Schülerbeurteilung setzt förderdiagnostische Kompetenzen der Lehrkräfte voraus. Die Gefahr, dass stereotype Äußerungen verwendet werden, ist groß, vor allem weil das Schreiben der verbalen Beurteilungen, im Gegensatz zur Ziffernnotenvergabe, viel Zeit erfordert (Freese, 1990; Valtin et al., 2002). Black und Wiliam (2010) sehen Lernen als einen interaktiven Prozess zwischen Schüler\*in und Lehrer\*in, die lehrende Person muss über den aktuellen Leistungsstand und Probleme, die der Potenzialentfaltung den/der jeweiligen Schüler\*in im Weg stehen, Bescheid wissen. Es geht dabei um komplexe soziale, emotionale und motivationale Prozesse, die es zu überblicken gilt, um die Schüler\*innen schlussendlich zu besseren Lernenden zu machen und auch ein erfolgreiches lebenslanges Lernen zu unterstützen. Es gibt viele fundierte Nachweise, die für einen Zusammenhang von formativer Leistungsbeurteilung (Feedback zur Anpassung von Lehr-Lern-Verhalten) im Unterricht und besser erzielten Leistungen der Schüler\*innen sprechen.

Ein formatives verbales Feedback hilft schwächeren Schüler\*innen mit und ohne Lernleistungsbeeinträchtigungen, die Differenzen zu den besseren Schüler\*innen auszugleichen, somit wird die Leistung der unterrichteten Klasse homogener, was wiederum positive kumulative Effekte für den gesamten Klassenverband haben könnte (Fuchs et al., 1997).

Die TIMSS (1995, Third International Mathematics and Science Study) war die erste groß angelegte Studie mit direktem Einblick in die Lehr-Lern-Prozesse zwischen Lehrer\*innen und Schüler\*innen. Es wurde der Mathematikunterricht in 41 Ländern analysiert und verglichen. Stigler und Hiebert (1997) verglichen basierend auf den gesammelten Daten der TIMSS achte Schulklassen der Länder Deutschland, Japan und der USA, weil diese als wichtige, ökonomisch gut situierte Länder gesehen wurden. Der

Vergleich mit Japan war von großem Interesse, weil die japanischen Schüler\*innen zum wiederholten Mal nahezu die Maximalpunktzahl im internationalen Vergleich der erbrachten Mathematikleistungen erzielten – Laut PISA-Studie (OECD, 2019) erzielte Japan im internationalen Vergleich mit 37 Ländern sogar den ersten Platz betreffend der Mathematischen Kompetenz der 15-jährigen Schüler\*innen. Stigler und Hiebert (1997) verglichen insgesamt 231 Klassen (100 in Deutschland, 50 in Japan, 81 in den USA). Deutsche und amerikanische Schüler\*innen verbrachten ihre Unterrichtszeit fast gänzlich damit, Routine-Prozeduren zu üben. Der japanische Unterricht beinhaltete komplexe Mathematik, einen klaren Fokus auf Denken und Problemlösen sowie eigenständige Suche nach Alternativlösungen und Verbalisierung des Denkens. Basierend auf den von den Schüler\*innen gegebenen Vorschlägen wurde von den Lehrkräften aufgabenzentriertes und individuelles Feedback gegeben, wohingegen in den amerikanischen und deutschen Schulen vorwiegend weniger differenzierte Rückmeldungen seitens der Lehrpersonen durchgeführt wurden. Feedback muss nachvollziehbar und aufgabenorientiert gegeben werden, nur so können Schüler\*innen ihr Lernverhalten an die gestellten Anforderungen anpassen (Noor, et al. 2010).

Im internationalen Vergleich schnitten auch skandinavische Länder mit Schulsystemen, die den Fokus auf individuelles Feedback und Verzicht auf Ziffernnoten bis zur achten Schulstufe legten, wie Finnland und Schweden, überdurchschnittlich gut in den PISA-Leistungstests ab (Baumert et al., 2001; OECD, 2019).

Im NOVARA-Projekt (Noten- oder Verbalbeurteilungen – Akzeptanz, Realisierung und Auswirkungen in Ost- und Westberliner Grundschulen) von Valtin et al. (2002) wurde im Längsschnittdesign untersucht, ob sich verbale Leistungsbeurteilungen positiver auf die schulischen Selbstkonzepte von Schüler\*innen auswirken als Notengebung. Ein Teil der Stichprobe erhielt Noten und Ziffernzeugnisse, während der andere Teil Verbalbeurteilungen und Berichtszeugnisse erhielt. Die untersuchte Lernfreude, bezogen auf das Fach Mathematik, blieb bei Kindern mit Verbalbeurteilungen gleich, wogegen sie bei Schüler\*innen mit Notenbeurteilungen von der 3. zur 4. Schulklasse abnahm (Valtin et al., 2002). Die Lernfreude vor allem von Buben sank beim Übergang

von der Primar- zur Sekundärschule. Bei den Mädchen ist ein signifikanter Abfall erst am Beginn der Sekundarstufe festzumachen (Valtin et al., 2002; Schmude, 2005). Die Selbstkonzeptveränderungen beider Gruppen waren gleich, der erwartete Vorteil von Verbalbeurteilungen gegenüber Notenbeurteilungen blieb in dieser Untersuchung aus (Zeinz & Koller, 2006).

Welche Auswirkung Leistungsrückmeldungen auf die Lernleistung und -motivation von Schüler\*innen haben, hängt von verschiedenen Faktoren ab. Dabei spielen sowohl Merkmale der/des Lehrenden, der/des Lernenden als auch die Aufgabe eine Rolle. Je nach Aufgabenschwierigkeit und dem vorhandenen Fachwissen des Kindes sind entweder elaborierte Rückmeldungen oder kurze „richtig/falsch“-Angaben effektiver (Rakoczy et al., 2019; Ruiz-Primo & Brookhart, 2017). Die positiven Effekte verbaler Beurteilungen erfordern eine angepasste fachdidaktische Lehrkräfteausbildung. Ohne adäquate Kompetenzen der Lehrer\*innen ist eine empirisch belegbare Verbesserung der leistungsbezogenen Fähigkeiten sowie des Selbstkonzeptes der Schüler\*innen nur schwer nachvollziehbar (Schmude, 2005). Laut Voerman et al. (2012) geben Lehrpersonen auf erbrachte Leistungen der Schüler\*innen vorwiegend undifferenzierte und kaum den Lernprozess förderliche Rückmeldungen. Deshalb sollte verbale Leistungsbeurteilung als zentraler Gegenstand der Lehrer\*innenausbildung betrachtet werden, damit die motivationsförderlichen Stärken der verbalen Leistungsbeurteilung zur Geltung kommen können (Hess, 2019; Wiliam, 2014). In weiterer Folge sind Schüler\*innen, die ihre eigenen Fähigkeiten gut einschätzen und Leistungsrückmeldungen nachvollziehen sowie anwenden können, erfolgreichere Lernende (Carless & Boud, 2018).

Auch trotz guter Ausbildung der Lehrpersonen kann es im Prozess der Leistungsbeurteilung zu Fehlern und „Verzerrungen“, kommen. Im folgenden Kapitel wird näher darauf eingegangen.

## 2.4. Beobachtungs- und Beurteilungsfehler bei der Leistungsbeurteilung

Beobachtungsfehler im Rahmen einer Leistungsbewertung können auftreten, wenn die Beurteilung einer Lehrkraft aufgrund unzureichender Sorgfalt, Ermüdung oder Langeweile beeinflusst wird. Diese können die Objektivität eines Resultates verringern und zu einer ungenauen oder falschen Beurteilung führen. Aber auch trotz einer fehlerfreien Beobachtung der für die Beurteilung relevanten Leistungen, können unterschiedliche Beurteilungsfehler auftreten (Häcker et al., 2013).

Lehrkräfte bewerten Schülerleistungen aufgrund von Verzerrungseffekten und anderen Fehlerursachen (z.B. Stereotype) unterschiedlich (Spiel, 1997). Bspw. tendieren Lehrpersonen dazu, Schüler\*innen mit niedrigem sozio-ökonomischen Status geringere Leistungserwartungen entgegenzubringen (Figlio, 2005) und Schüler\*innen mit Migrationshintergrund schlechter zu benoten (Sprietsma, 2013). McKown und Weinstein (2003) gingen aufgrund der Ergebnisse ihrer Untersuchung davon aus, dass Schüler\*innen aus stigmatisierten Gruppen, mit Bewusstheit über die Stereotype anderer, in diagnostischen Situationen schlechter abschneiden als Kinder, die keinen stigmatisierten Gruppen angehören (vgl. stereotype threat effect; Spencer et al., 2016).

Ziffernnotenbeurteilungen sind naturgemäß aufgrund der Orientierung an der sozialen Bezugsnorm verzerrt (Schmid et al., 2016). In Relation zur durchschnittlichen Klassenleistungsfähigkeit erhalten Schüler\*innen von Grundschulen schwerer oder leichter eine bessere Leistungsbeurteilung. Der *Big-fish-little-pond-effect* (dt. Fischteicheffekt, Marsh, 1987; Marsh et al., 2014) beschreibt, dass zwei Schüler\*innen in unterschiedlichen Schulklassen trotz identischer erbrachter Leistungen verschiedene Benotungen erhalten und deshalb unterschiedliche akademische Selbstkonzepte entwickeln können. Somit profitieren leistungsstärkere Schüler\*innen derselben Klasse mit ihren akademischen Selbstkonzepten auf Kosten der schwächeren (Zeinz & Köller, 2006). Dieses Phänomen kommt vor allem in Klassen mit Notenbeurteilungen vor und

ist in Schulsystemen mit früher Selektion im Gegensatz zu Gesamtschulsystemen ausgeprägter (Salchegger, 2016).

Abgesehen von den auf der sozialen Bezugsnorm basierenden Fehltritten, kann es zu weiteren Beurteilungsfehlern kommen (Häcker et al., 2013): Der *Milde-Effekt* trägt zu einer zu positiven Beurteilung aufgrund von Voreingenommenheit bei; *Großzügigkeitsfehler* tragen, wegen des verlorenen Bezugs zur sachlichen Norm, zu einer tendenziell zu positiven Beurteilung bei; *Tendenz zur Mitte* (extreme Urteile werden gemieden) und *Tendenz zu extremen Urteilen* (mittlere Resultate werden gemieden) sowie *Halo-Effekte* können aufgrund markanter Merkmale (bspw. äußeres Erscheinungsbild oder Mundart) die Bewertung beeinflussen (Häcker et al., 2013).

### ***Der Halo-Effekt im schulischen Kontext***

Thorndike (1920) postulierte den *Halo-Effekt* als einen systematischen Fehler der Personenbeurteilung, bei dem ein einzelnes zu bewertendes Merkmal nicht objektiv von Einflüssen anderer für die Messung irrelevanter Merkmale getrennt beurteilt werden kann. Dieses Phänomen tritt sowohl bei Laien als auch bei Expert\*innen, wie bspw. Lehrkräften, auf.

Laut Nisbett und Wilson (1977) kann ein *Halo-Effekt* auch auftreten, wenn für die Beurteilung der notwendigen Fragestellung vollkommen ausreichende Informationen vorhanden sind. In ihrer Studie beurteilten Proband\*innen auf Video aufgenommene Interviews hinsichtlich des Auftretens eines englisch sprechenden Lehrers mit europäischem Akzent. In einem der Videos verhielt sich der Lehrer warmherzig und freundlich, im anderen kalt und distanziert. Proband\*innen bewerteten den Akzent des Lehrers im freundlichen Durchgang als ansprechend. Im Gegensatz dazu wurde sein Akzent bei distanzierterem Auftreten als irritierend empfunden. Die Teilnehmer\*innen bewerteten also die Eigenschaften des kaltherzig wirkenden Lehrers als negativer, gaben aber an, dass die Abneigung gegenüber dem Lehrer keinen Einfluss auf die Beurteilung nahm. Nisbett und Wilson (1977) beschrieben diese Fehlbeurteilung als eine fundamentale Unfähigkeit, spezifische Beurteilungen ohne den Einbezug von

emotionalen Einflüssen globaler Bewertungen durchzuführen. Cooper (1981) nennt unter anderem zugrundeliegende illusorische Theorien der Beurteiler\*innen, die die kognitiven Verzerrungen hervorrufen. Sowohl Charakteristika der Beurteilungssituation sowie des/der Beurteilers/in selbst beeinflussen den *Halo-Effekt*. So nehmen Murphy und Anhalt (1992) an, dass der *Halo-Effekt* ein Ergebnis zugrundeliegender kognitiver Schemata und impliziter Persönlichkeitstheorien der/des Beurteilers/in ist.

Eine Meta-Analyse von 20 experimentellen Studien mit insgesamt 1935 Beurteiler\*innen wurde von Malouff und Thorsteinsson (2016) durchgeführt. Die Studien befassten sich mit Urteilsverzerrungen betreffend die Benotung von Schüler\*innen und Studierenden, u.a. beim Schreiben von Aufsätzen. Sie untersuchten implizite bzw. unbewusste Beeinflussungen der Leistungsbeurteilung durch spezifische Nebeninformationen der hinter den zu beurteilenden Arbeiten stehenden Personen. Dazu gehörten Informationen zu ethnischen Hintergründen, Bildungsdefizite, physische Unattraktivität und vorangegangene erbrachte schlechte Leistungen. Die meta-analytische statistisch signifikante Effektstärke zwischen den Testgruppen und den Kontrollgruppen betrug  $g = 0.36$ , also eine Standardabweichung von 36 % [KI: .24, .48], was für einen kleinen bis mittleren Verzerrungseffekt bei den Beurteilungen spricht (Cohen, 1992). Sämtliche oben genannten Vorinformationen wirkten sich signifikant auf die Beurteilung aus. Dieser Effekt zeigte sich unabhängig davon, ob es sich bei den benoteten Personen um Grundschüler\*innen, Sekundarschüler\*innen oder Studierende handelte. Durch diese Meta-Analyse zeigt sich eine hohe Anfälligkeit der zu beurteilenden Personen durch irrelevante Nebeninformationen über die Schüler\*innen und Studierenden (Malouff & Thorsteinsson, 2106). So kommt es bei einem generell positiven Eindruck eines/einer Schülers\*in auch ohne die erforderlichen Nachweise zu einer zu positiven Beurteilung der jeweiligen erbrachten Leistung (Talamas et al., 2016).

Der durch den *Halo-Effekt* herbeigeführte systematische Urteilsfehler verringert die Validität der jeweiligen Beurteilung und sorgt somit für Probleme bei Leistungsbeurteilungen und -vergleichen in unterschiedlichen Selektionsverfahren, wie bspw. Bewerbungsgesprächen und Aufnahmetests (Feeley, 2002).

Beurteiler\*innen-Trainings (Hoyt & Kerns, 1999), kombinierte Bewertungen mehrerer Beurteiler\*innen und Anonymität der zu beurteilenden Person (Brennan, 2008) können den *Halo-Effekt* reduzieren. International betrachtet nimmt Österreich eine Randposition ein, weil an den Schulen nahezu sämtliche Prüfungen – ausgenommen bspw. Zentralmatura und kommissionelle Prüfungen – von denselben Personen beurteilt werden, die die Prüfungsinhalte auch vorab gelehrt haben. Für österreichische Schulen sind deshalb die oben genannten Möglichkeiten der Beurteilungsfehlerreduktion in dem bestehenden Bildungssystem nicht anwendbar (Eder et al., 2009).

## 2.5. *Offenheit (Big-Five)* bei Lehrer\*innen

Der Psychologe Hans Jürgen Eysenck (1947) gilt als Pionier der fünf grundlegenden Faktoren menschlicher Persönlichkeitseigenschaften: *Extraversion*, *Neurotizismus*, *Gewissenhaftigkeit*, *Offenheit* und *Verträglichkeit* (Cattell, 1957; Cattell, 1965). Goldberg (1981) prägte den Begriff „Big-Five“. Die faktorenanalytisch gewonnenen fünf Dimensionen wurden seither mittels vielerlei entwickelter Tests (u.a. Satow, 2012) für Forschung und praxisrelevante Fragestellungen, wie bspw. Ausbildungs- und Job-Qualifikationen (Schmidt & Hunter, 1998; Schuler, 2005), herangezogen.

Personen mit hohen *Offenheit*-Werten sind besonders tolerant, neugierig und wissbegierig (Satow, 2011). Anpassungsfähigkeit und eine erhöhte Suche nach unbekanntem, neuen Möglichkeiten im Umgang mit der Umwelt, sind bei Menschen mit hohen Werten in *Offenheit* auch eher gegeben (McCrae & Costa, 1997).

Fayn et al. (2017) versuchten, zwei unterschiedliche Positionen zur Unterscheidung des inhaltlichen und affektiven Engagements zwischen *Offenheit* und Intellekt durch Reaktionen und Bewertungen auf eine breite Palette von visuellen und abstrakten Reizen zu testen. 191 Proband\*innen nahmen an der Untersuchung teil. Ihre Aufgabe war es, Bilder von Kunst, philosophischen Zitaten und wissenschaftlichen Entdeckungen anzusehen und hinsichtlich der Kategorien Interesse, Neuheit und Verständnis zu bewerten. Im Gegensatz zu Intellekt konnten bei *Offenheit* in allen drei Kategorien Zusammenhänge gefunden werden.

Eder (2008) untersuchte 458 Studienanfänger\*innen für das Lehramtsstudium hinsichtlich der *Big-Five*-Persönlichkeitseigenschaften. Im Vergleich zur Normstichprobe zeigten sich überdurchschnittliche Ausprägungen in sämtlichen Persönlichkeitsmerkmalen. Auch Mayr (2009) kam zu dem Ergebnis, dass die *Big-Five* bei Lehrer\*innen stärker ausgeprägt seien, wobei der Faktor *Offenheit* bei Lehramtsstudierenden im Gegensatz zur untersuchten Referenzstichprobe die größten Unterschiede ausmachte.

Eulenberger (2015) bestätigte in seiner Untersuchung (VEBOLAS: Verbleib und berufliche Orientierung von sächsischen Lehramtsabsolvent\*innen in Sachsen) die dominante Ausprägung der Persönlichkeitseigenschaft *Offenheit* bei Lehramtsabsolvent\*innen, Lehramtsstudierenden und Lehrer\*innen im Vergleich zum Bevölkerungsdurchschnitt.

In der Studie von Mohan und Mulla (2013) zeigten sich Ergebnisse, dass sich *Offenheit* negativ auf die Leistung bei Jobs mit geringer Komplexität und positiv bei Jobs mit hoher Komplexität auswirkt. Sie heben in ihren Ergebnissen hervor, dass die Beziehung zwischen *Offenheit* und Arbeitsleistung von den beruflichen Rahmenbedingungen abhängig ist. *Offenheit* sei in ihrem Einfluss tendenziell bedeutender, wenn der berufliche Kontext genügend Autonomie bietet, unkonventionelles Denken erfordert und individuelle Freiheit schätzt.

### ***Offenheit und Leistungsbeurteilung***

Eine im deutschsprachigen Raum durchgeführte Untersuchung von 648 Lehramtsstudierenden an pädagogischen Akademien kam zu dem Ergebnis, dass sich Studierende mit höheren Werten in *Offenheit* „aktiv-kritischer“ mit ihren Studieninhalten beschäftigten (Mayr, 2003). Faust und Foerster (2008) waren der Meinung, „offenere“ Lehrkräfte seien eher dazu bereit, heterogene Klassen mit angepassten Lehrmethoden zu unterrichten, um den unterschiedlichen Voraussetzungen der Schüler\*innen gerecht zu werden. Eine Begründung für diese Annahmen könnte mit dem Konstrukt der „Ambiguitätstoleranz“ (Hartinger et al., 2005) zusammenhängen. Nach diesem Konstrukt streben „ungewissheitstolerante“ Personen neue und zu bewältigende Herausforderungen an, wohingegen „ungewissheitsintolerante“ Personen eher dazu neigen, diese als Bedrohung anzusehen (Sorrentino & Roney, 2000).

Ein Beispiel dafür ist die Studie von Camadan et al. (2018), in der ein Zusammenhang von Persönlichkeitseigenschaften und der Bereitschaft, sich mit neuen Technologien auseinanderzusetzen, untersucht wurde. Für diese Studie erhielten Lehrer\*innen neue, zur Verwendung im Unterricht gedachte Tablet-PCs und ein Training für den Umgang

mit den Geräten. Personen mit höherer *Offenheit* waren eher davon überzeugt, mit den neuen Geräten produktiv und effektiv umgehen zu können. Es zeigte sich außerdem, dass Personen mit hohen Werten in *Offenheit* eher dazu neigen, einen positiven Nutzen in neuen Technologien zu sehen und eine erhöhte Bereitschaft aufweisen, sich darauf einzulassen.

Die Ergebnisse der Quantifizierung von 15 metaanalytischen Studien von Barrick et al. (2001) ließen auf Zusammenhänge von *Offenheit* mit Intellekt, Neugierde, Vorstellungsvermögen und Trainingskompetenz schließen. Personen mit hoher *Offenheit* waren eher bereit, sich auf Arbeitstrainings, bzw. das Lernen neuer arbeitsrelevanter Inhalte einzulassen. Ebenso konnten Nachweise für höhere Trainingsleistungen und *Offenheit* erbracht werden. Die voneinander unabhängigen Untersuchungen erbrachten moderate Effekte zwischen Trainingskompetenz und *Offenheit*.

Tan et al. (2019) untersuchten die Beziehung zwischen *Offenheit* und Kreativität. Die Definition für Kreativität variiert in der Literatur, Tan et al. (2019) bedienten sich der Begriffsbestimmung für Kreativität von Erez & Nouri (2010), wobei diese mit Originalität und Nützlichkeit beschrieben wird. Ein Ergebnis (Produkt oder Idee) einer kreativen Handlung gilt als an die jeweilige Fragestellung angepasst und einzigartig. *Offenheit* wurde als guter Prädiktor für die Zeit, die für kreative Aufgaben aufgewendet wird, erkannt (Silvia et al., 2014). Vorschullehrer\*innen zeigten innerhalb der zwei Persönlichkeitsmodelle Eysenck (1967) und *Big-Five* (Cattell, 1957) höhere Werte in *Extraversion, Verträglichkeit, Gewissenhaftigkeit, Offenheit* und sozialer Konformität gegenüber der Referenzpopulation. Die in Anbetracht der Referenzpopulation höheren Werte in *Offenheit* könnten auf eine wesentliche Kompetenz der Vorschullehrer\*innen hindeuten. Eine der wichtigsten Rollen besteht darin, sich im Sinne der pädagogischen Arbeit auf die Schüler\*innen einzulassen, um sie im schulischen Rahmen aus verschiedenen Blickwinkeln betrachten und beurteilen zu können (Tatalović Vorkapić, 2012).

## 2.6. Beurteilung durch Ziffernnoten versus Verbalbeurteilung

Im folgenden Abschnitt wird ein Zwischenfazit aus der behandelten Literatur gezogen und anschließend empirische Studien zu den Präferenzen der Beurteilungsmethoden von Eltern, Lehrer\*innen und Schüler\*innen dargestellt.

### **Zwischenfazit**

Einerseits besteht durch die Beurteilung mithilfe von Ziffernnoten die Möglichkeit einfacher interindividueller Vergleiche von Schüler\*innen, andererseits erhöhen diese Vergleichsmöglichkeiten den auf die Kinder ausgeübten Druck. Weil Ziffernnotenzeugnisse über den Aufstieg in die Sekundarstufe entscheiden, spielen die Allokations- und Qualifikationsaspekte der Leistungsbeurteilungen eine dominante Rolle für das Vorankommen in unserem Bildungssystem (Spiel & Wagner, 2003). Das selektionsorientierte Ziffernzensursystem unterstützt Konkurrenzdenken sowie Prüfungsangst und hemmt gleichzeitig Kompetenzen im Umgang mit Konflikten sowie Kooperationsbereitschaft unter Mitschüler\*innen. Der Fokus der Schüler\*innen wird von den eigentlich zu lernenden Inhalten weg zu der Fragestellung, wie es am leichtesten möglich ist, gute Noten zu erhalten, gelenkt (Eccles, 2005; Meece et al., 2006). Die pädagogischen Notwendigkeiten, differenziertes Feedback über die erbrachte Leistung, Lernstrategien und Fähigkeiten der Schüler\*innen zu liefern, um die jeweiligen Fähigkeiten zu verbessern, kann durch die Vergabe von Ziffernnoten also nicht erfüllt werden (Ingenkamp, 1995; Spiel & Wagner, 2003).

Aufgrund der sozialen Bezugsnorm (vgl. *big-fish-little-pond-effect*, Marsh et al., 2014) und des möglichen Einflusses irrelevanter Nebeninformationen des eigentlich zu beurteilenden Lernstandes (*Halo-Effekt*, Malouff & Thorsteinsson, 2016) stellt sich die Leistungsbeurteilung durch Ziffernnoten als besonders fehleranfällig dar. Insgesamt lassen Ziffernnoten nur sehr eingeschränkt Rückschlüsse auf die Fertigkeiten von Schüler\*innen in den verschiedenen Schulfächern zu (OECD, 2019; Paasch et al., 2019).

Schüler\*innen werden durch die Art des verwendeten Leistungsfeedbacks maßgebend beeinflusst. Wenn die Klassenkultur den Fokus auf das Verteilen von „goldenen Sternchen“, Ziffernnoten oder Rangreihungen legt, werden Schüler\*innen versuchen, bessere Beurteilungen zu bekommen, anstatt bessere Lernleistungen zu erzielen. Dies könnte ebenso zu Leistungsangst und dem Attribuieren von Fehlleistungen zur eigenen Persönlichkeit sowie verringerten Selbstwirksamkeitserwartungen führen (Black & Wiliam, 2010). Die verbalen Beurteilungen sollen als integrierter Aspekt im Unterrichten verflochten sein und nicht (nur) in gesonderten Prüfungen stattfinden (Vierlinger, 2012). Verbalbeurteilung kann ein nützliches Werkzeug sein, um negative Effekte, wie die eben genannten und den Wettkampf zwischen den Schüler\*innen zu reduzieren und Inhaltsverständnis sowie Qualität ihrer Leistungen zu verbessern. Rückmeldungen für Schüler\*innen sollten sich hauptsächlich auf die individuellen erbrachten Leistungen und Möglichkeiten der Verbesserung dieser fokussieren (Black & Wiliam, 2010).

Ein Vorteil der verbalen Leistungsbeurteilungen ist die Möglichkeit präziser Rückmeldungen über die kognitiven Fähigkeiten der Schüler\*innen an die jeweiligen Eltern. Vor allem Kinder aus Familien mit niedrigem sozio-ökonomischem Status würden mit erhöhten Bildungschancen von einem Schulsystem profitieren, in dem eine Notenbeurteilung erst später durchgeführt wird (Filippin & Paccagnella 2012; Hastings & Tejada-Ashton, 2008).

Das pädagogische Konzept von Leistung und Lehrerhandeln wirkt sich auf die Lernfreude der Schüler\*innen aus. Ein begreiflicher, abwechslungsreich gestalteter Unterricht sowie die Möglichkeit für die Schüler\*innen, auch selbstständige Entscheidungen zu treffen, sind prägend für das Ausmaß der Lernfreude von Schüler\*innen. Diese Bedingungen erfordern eine reflektierte und differenzierte Form der Leistungsrückmeldung bzw. -beurteilung, die mithilfe von Verbalbeurteilungen stattfinden kann. Ein ausgeprägtes Ausmaß an *Offenheit* wurde in der Literatur bereits als nützlich im Rahmen des Lehrberufes beschrieben. So wird die bei Lehrer\*innen höhere *Offenheit* (Eulenberger, 2015) mit erfolgreicher pädagogischer Arbeit mit den individuellen Bedürfnissen ihrer Schüler\*innen in Verbindung gebracht (Tatalović

Vorkapić, 2012). Empirische Vergleiche setzen aber voraus, dass erstens die Lehrkraftausbildung die nötigen psychologischen Kompetenzen beinhaltet und zweitens Verbalbeurteilungen in den jeweiligen zu untersuchenden Schulsystemen bereits über eine längere Zeit durchgeführt werden (Helmke & Weinert, 1997; Helmke & Weinert, 1999; Schneider & Bullock, 2008).

Weiters sind die fehlenden Vergleichsmöglichkeiten von Verbalbeurteilungen im Gegensatz zu den Ziffernnoten in unserem Schulsystem problematisch. Die Allokation, also die Berechtigung zum Aufstieg in die nächste Schulstufe auf Basis des Notendurchschnitts, nimmt in unserem Bildungssystem einen hohen Stellenwert ein, denn gute Noten in der vierten Klasse sind die Voraussetzung für die Berechtigung zum Aufstieg in die AHS (Spiel & Wagner, 2003). In Deutsch und Mathematik dürfen die Noten nicht schlechter als „Gut“ sein und die restlichen Pflichtgegenstände müssen zumindest positiv bewertet werden (LBVO, 1974; Paasch et al., 2019).

Solange die Schulnoten über den Aufstieg in die AHS bestimmen, ist die Leistungsbeurteilung durch Ziffernnoten ein fester Bestandteil der Primarstufe (Paasch et al., 2019).

## ***Empirische Studien zur Präferenz der Leistungsbeurteilung: Beurteilung durch Ziffernnoten versus Verbalbeurteilungen***

In den folgenden zwei empirischen Studien werden die Ergebnisse hinsichtlich der Präferenz für die Beurteilungsmethoden (Ziffernnoten oder Verbalbeurteilung) hervorgehoben.

### **NOVARA (Noten- oder Verbalbeurteilung? Akzeptanz, Realisierung und Auswirkungen in Ost- und Westberliner Grundschulen)**

Valtin und Rosenfeld (2002; 1997) untersuchten in ihrem Forschungsprojekt unter anderem die Akzeptanz gegenüber der seit 1991 eingeführten Berichtszeugnissen, die die bisherigen Notenzeugnisse an Berliner Grundschulen ersetzen. Evaluiert wurden die Präferenzen für Beurteilungen durch Ziffernnoten und Verbalbeurteilungen von West- und Ostberliner Eltern und Lehrer\*innen. Die Untersuchung startete im Schuljahr 1994/95 mit 434 Schüler\*innen aus 23 Berliner Grundschulen. Schlussendlich nahmen davon 95 Schüler\*innen bis zum Abschluss der Grundschule an der Erhebung teil. Valtin und Rosenfeld (1997) weisen in ihrem Forschungsbericht darauf hin, dass mehr Schulklassen mit Verbalbeurteilungen in die Untersuchung einbezogen wurden und somit eine Verzerrung der Ergebnisse nicht auszuschließen ist.

Eltern aus Westberlin (n = 212) gaben Großteils an, sich der negativen Auswirkungen von Schulnoten (bspw. Schulangst und Konkurrenzdenken) und der positiven Möglichkeiten von Verbalbeurteilungen (detaillierte Lernstandsanalyse für zielgerichtete Fördermöglichkeiten) bewusst zu sein. Im Gegensatz dazu, schrieben die Eltern aus Ostberlin (n = 219) den Beurteilungen durch Ziffernnoten positive Aspekte zu („Objektivität und Aussagekraft“).

Auf die Frage hin, in welcher Klasse Ihre Kinder welche Form der Leistungsbeurteilung erhalten sollten, wurde folgend geantwortet: Die „Oststichprobe“ gab mehrheitlich an, nur im ersten Schuljahr eine verbale Leistungsbeurteilung zu bevorzugen (67.6 %). Bei der „Weststichprobe“ gab es bei der ersten (78.3 %) und der zweiten Schulklasse (55.0 %) eine Mehrheit für die Verbalbeurteilung der Kinder.

Nur ein Anteil von ungefähr 2.5 % der gesamten „Elternstichprobe“ wollte eine Rückführung zu Notenzeugnissen ohne Beurteilungstexte. Valtin und Rosenfeld (2002) stellten im Zuge Ihrer Untersuchung bei den Westberliner Eltern Zusammenhänge zwischen der Präferenz für Verbalbeurteilungen sowie Ablehnung von Ziffernnoten und zeitgenössischer öffentlicher Diskussion der Notengebung hinsichtlich pädagogischer, wissenschaftlicher Aspekte fest. Sie vermuteten auch, der positive Zugang zu Schulnoten der Ostberliner Eltern könnte aufgrund der damaligen „freundlicheren“ Benotungen und deshalb geringen Anzahl an „Sitzenbleibern“ für die Kinder weniger bedrohlich gewirkt haben (Valtin & Rosenfeld, 2002; 1997).

### **LeiHS (Leistungsbeurteilung und Leistungsrückmeldung an Hamburger Schulen)**

Lütgert et al. (2001) führten im Zeitraum von Juli 1997 bis Dezember 1999 das Projekt „Leistungsbewertung und -rückmeldung an Hamburger Schulen“ (LeiHS) durch. Im Fokus dieser empirischen Untersuchung standen Ziffernzeugnisse, Lernentwicklungsberichte sowie Mischformen beider Leistungsbeurteilungsmethoden von Schulen der Primar- sowie Sekundarstufe. Eltern, Lehrer\*innen und Sekundarschüler\*innen wurden mithilfe von standardisierten Fragebögen und 61 Grundschüler\*innen wurden, aufgrund der Komplexität der Fragebögen, mittels qualitativen Interviews befragt. Zehn Grundschulen nahmen an der Untersuchung teil, 146 Fragebögen von Lehrer\*innen wurden ausgewertet und 312 Eltern wurden von zwei dieser Grundschulen befragt. Folgend wird näher auf die von den Proband\*innen erwähnten Vor- und Nachteile der verschiedenen Beurteilungsformen sowie die daraus resultierenden Präferenzen eingegangen.

Nur ungefähr 18 % der befragten Lehrer\*innen sprachen sich grundsätzlich gegen Zeugnisse und Benotungen aus. Die Beurteilungsfunktion der Schule wurde von den meisten Lehrpersonen als wichtig angesehen. Ein Drittel der Grundschullehrkräfte gab an, für eine Schule ohne Noten zu sein. Diese Einstellung verändert sich in der Sekundarstufe drastisch, nur 4 % der Gymnasiallehrkräfte waren dieser Meinung – was

mit der fehlenden Praxis im Umgang mit Verbalbeurteilungen zusammenhängen könnte, weil in der Sekundarstufe keine Berichtszeugnisse ausgestellt werden. Hier ist anzumerken, dass Grundschullehrer\*innen in der Regel Praxiserfahrung mit beiden Beurteilungs- und Zeugnisformen haben.

Die befragten Lehrer\*innen sprachen sich, im Vergleich zu den Zensurzeugnissen, hauptsächlich für Berichtszeugnisse und deren pädagogischen Nutzen, aus. Je nach Schulform gab es Unterschiede in den Präferenzen der Lehrer\*innen. Bspw. waren Grundschullehrkräfte zu einem hohen Anteil für Berichtszeugnisse wohingegen Gymnasiallehrer\*innen für Notenzeugnisse plädierten. Trotz dieser Differenzen gab es einen Konsens, und zwar „Notenzeugnisse mit Kommentarbögen“ wurden von allen Lehrpersonen hauptsächlich positiv bewertet. Zusätzlich zu der Schulform wurden die Lehrer\*innen durch Dienstalter und die Erfahrung mit bestimmten Beurteilungsformen hinsichtlich ihrer Präferenzen beeinflusst. Ein höheres Dienstalter sprach eher für eine Notenbevorzugung und je mehr Erfahrung mit einer Beurteilungsform, desto eher wurde sie auch bevorzugt.

Ähnliche Unterschiede hinsichtlich der Schulformen gab es auch bei den Eltern. Je nach Zeugnisart der Schule der eigenen Kinder wurde diese jeweils positiver eingeschätzt. Weiters sprechen sich Eltern mit höherem Bildungsabschluss meist eher für Berichtszeugnisse aus.

Die meisten Grundschüler\*innen interessieren sich für ihre Berichtszeugnisse und können sie als für sie maßgeschneiderte Informationen wahrnehmen, sofern sie beim Verstehen der Texte von Eltern und Lehrer\*innen unterstützt werden. Weshalb auch dem Lesen der Zeugnisse sowohl in der Schule als auch zu Hause genug Raum und Zeit gegeben werden sollte. Unter den befragten Schüler\*innen der Primarstufe findet sich eine Mehrheit, die sich für die Bevorzugung von verbalen Beurteilungen ausspricht. Auch wenn die Schüler\*innen keine eigenen Erfahrungen mit Beurteilungen durch Ziffernnoten machen konnten, sahen die meisten Zensurzeugnisse bereits bei Geschwistern oder Freund\*innen und konnten sich daher eine Meinung bilden. Aus verbalen Beurteilungen hervorgehende Kritik wurde von den Kindern meist als Ansporn

für weitere Anstrengungen wahrgenommen. Berichtszeugnisse werden als individualisierte Texte und weniger als Zeugnisse verstanden – was auch von den schwächeren Kindern verstanden werden kann.

Als Gemeinsamkeit kann die positive Wirkung von (guten) Zensuren und Verbalbeurteilungen auf die Lernmotivation hervorgehoben werden. Die mangelnden Gütekriterien von Ziffernnoten wurden bei Angaben der Präferenzen nicht näher berücksichtigt. Weiters ist die schulformspezifische Bevorzugung für die Leistungsbeurteilungsart auffällig. In der Primarstufe wird die verbale Leistungsbeurteilung von Eltern sowie Lehrer\*innen präferiert, spiegelbildlich stellen sich die Präferenzen von Eltern und Lehrer\*innen im Kontext der Sekundarstufe dar. In allen Gruppen (ausschließlich der Grundschüler\*innen) herrscht, auch wenn damit noch keine Erfahrungen gemacht wurden, zusätzlich eine Präferenz für Notenzeugnisse mit Kommentarbögen (Lütgert et al., 2001).

### 3. Fragestellung und Hypothesen

Im Schuljahr 2019/20 kam es österreichweit zur Wiedereinführung der Beurteilung durch Ziffernnoten von Volksschüler\*innen ab dem Jahreszeugnis der zweiten Klasse. Die Möglichkeit der verbalen Leistungsbeurteilung der Schüler\*innen ist weiterhin bis zum Ende des ersten Semesters der zweiten Schulklasse möglich, dafür ist jedoch eine Übereinkunft der jeweiligen Eltern, innerhalb der ersten neun Wochen des Schuljahres, notwendig (SchuG §18a). Es stellt sich die Frage nach der Sinnhaftigkeit des „Rückschrittes“ in der Leistungsbeurteilung von Volksschüler\*innen. Notenbeurteilungen weisen in sämtlichen Gütekriterien für Beurteilungsinstrumente - Objektivität, Validität und Reliabilität - Mängel auf (Valtin et al., 2002). Verbalbeurteilungen setzen Kompetenzen in der Durchführung voraus, um sinnvoll eingesetzt werden zu können (Black & Wiliam, 2010) und benötigen mehr Zeit, was zu stereotypen Aussagen führen kann (Valtin et al., 2002), haben aber das Potenzial den Lehr-Lern-Prozess positiv, durch lernleistungsförderliche und selbstbildstärkende Effekte, zu beeinflussen (Carless & Boud, 2018; Hess, 2019; Rakoczy et al., 2019; Wiliam, 2014).

Die Hauptforschungsfrage dieser Arbeit lautet:

*Welche Beurteilungsmethode wird nach erneuter Einführung der Beurteilung durch Ziffernnoten, ab dem Jahreszeugnis der zweiten Klasse der Volksschule, von Lehrpersonen und Laien je Schulklasse für die Schüler\*innen bevorzugt und ist die Beurteilung durch Ziffernnoten anfällig für Urteilsverzerrungen?*

Eine Kombination der beiden Beurteilungsmethoden, also gleichzeitige Noten- und Verbalbeurteilung, birgt die Gefahr der dominanten Einflussnahme der Ziffernnoten und somit der negativen Folgen der sozialen Bezugsnormorientierung, wie die wettbewerbsorientierten Vergleiche der Schüler\*innen untereinander und den Verlust der Lernzielorientierung (Black & Wiliam, 2010). Deshalb werden nur die zwei Methoden „Beurteilung durch Ziffernnoten“ und „Verbalbeurteilungen“ in dieser Untersuchung berücksichtigt. In der Untersuchung von Valtin und Rosenfeld (2002) kommt es zu Unterschieden in der mehrheitlichen Präferenz der Leistungsbeurteilung in der zweiten Klasse der Grundschule bei Eltern und Lehrer\*innen. Nachdem auch Eltern, bzw. Laien in die Entscheidung eingebunden werden, ob bis zum Ende des ersten Semesters der zweiten Klasse, Verbalbeurteilungen oder Schulnoten vergeben werden, werden diese auch in dieser Untersuchung miteinbezogen (SchuG §18a).

*Hypothese 1: Die Leistungsbeurteilungspräferenzen von Laien und Lehrer\*innen für Schüler\*innen in den vier Volksschulklassen unterscheiden sich.*

Die Persönlichkeitseigenschaft *Offenheit* (Eysenck, 1947) weist Zusammenhänge mit bevorzugt kreativen und zielgerichteten Vorgehensweisen auf (Erez & Nouri, 2010; Tan et al., 2019). Lehrkräfte mit ausgeprägter *Offenheit* weisen eine höhere *Ambiguitätstoleranz* auf – stehen also ungewohnten Herausforderungen eher positiv geneigt gegenüber – und seien eher dazu bereit, individualisierte Lehrmethoden für heterogene Klassen anzuwenden (Faust & Foerster, 2008). *Offenheit* sei im beruflichen Kontext von höherer Bedeutung, wenn autonomes, komplexes Handeln und unkonventionelles Denken erfordert werden (Mohan & Mulla, 2013; Tatalović Vorkapić, 2012).

*Hypothese 2: Es gibt einen positiven Zusammenhang zwischen Offenheit bei Laien und Lehrer\*innen und der Beurteilungspräferenz für Verbalbeurteilungen für Schüler\*innen in den vier Volksschulklassen.*

Ziffernnoten werden in der Literatur, vor allem wegen der zugrundeliegenden sozialen Bezugsnorm, als anfällig für Beurteilungsfehler beschrieben (Schmid et al., 2016; Spiel, 1997). Der *Halo-Effekt* (Thorndike, 1920) beeinflusst Beurteilungen einzelner Fähigkeiten und Kompetenzen einer Person, aufgrund irrelevanter, für den/die Beurteiler\*in salienter Wahrnehmungen. Diesen Wahrnehmungen liegen kognitive Schemata und implizite Persönlichkeitstheorien der Beurteilerin/des Beurteilers zugrunde (Anhalt, 1992; Cooper, 1981). Es wurden bereits viele Nachweise erbracht, dass Verzerrungseffekte, bspw. aufgrund von vorangegangenen schlechten Leistungen, oder eines generellen positiven Eindrucks des/der zu beurteilenden Schülers/Schülerin, bei Leistungsbeurteilungen auftreten können (Malouff & Thorsteinsson, 2016; Talamas et al., 2016).

*Hypothese 3: Lehrer\*innen werden bei der Beurteilung durch Ziffernnoten durch für die Beurteilung irrelevante Nebeninformationen beeinflusst.*

## 4. Methode

Im folgenden Teil werden das Studiendesign, die Stichprobe und die verwendeten Erhebungsinstrumente beschrieben.

### 4.1. Studiendesign

Bei der Untersuchung handelt es sich um eine Querschnittstudie mit einem Messzeitpunkt. Die zur statistischen Berechnung verwendeten sowie weitere explorativ-qualitative Daten wurden mittels einer Online-Erhebung im Zeitraum von Anfang November 2019 bis Ende Jänner 2020 über [www.soscisurvey.de](http://www.soscisurvey.de) durchgeführt. Die statistischen Analysen des erhobenen Datensatzes wurden mittels IBM SPSS Statistics v.25 berechnet.

Die Rekrutierung von Proband\*innen erfolgte bei Lehrer\*innen und Laien unterschiedlich. Lehrpersonen wurden mithilfe der österreichischen Schuldatenbank unter [www.schulen-online.at](http://www.schulen-online.at) kontaktiert. Dafür wurden insgesamt 237 Volksschulen mit der Bitte, an alle Mitarbeiter\*innen der jeweiligen Schulen, zur Unterstützung der Studie via E-Mail kontaktiert. Laien wurden über diverse Online-Foren (u.a. [www.elternforum.at](http://www.elternforum.at)) mit der Bitte zum Ausfüllen des Online-Fragebogens kontaktiert. Alle Proband\*innen nahmen freiwillig an der Untersuchung teil.

Die Proband\*innen wurden über die Tatsache aufgeklärt, dass die Studie im Rahmen einer Masterarbeit im Fach Psychologie am Institut für Psychologie der Entwicklung und Bildung an der Universität Wien durchgeführt wird.

## 4.2. Erhebungsinstrumente

### ***Beurteilungspräferenz***

Die Beurteilungspräferenz, also die Bevorzugung der Beurteilung durch Ziffernnoten oder Verbalbeurteilung je Volksschulklasse, wurde im Online-Fragebogen für jede der vier Schulklassen einzeln, sowohl von Lehrpersonen als auch Laien erfragt.

Das Item zum Erfragen der Beurteilungspräferenz je Schulklasse (vier Klassen) war:

„Geben Sie an, ob Ihrer Meinung nach Schulleistungen von Volksschüler\*innen mittels Verbalbeurteilungen oder Ziffernnoten von 1 bis 5 angemessener beurteilt werden können.“

Die Antwortmöglichkeiten je Schulklasse waren:

- 1.) Verbalbeurteilung
- 2.) Ziffernnoten
- 3.) Kann ich nicht beurteilen

Weiters bestand die Möglichkeit einer Erklärung der Auswahl mit einer freien Texteingabe.

## ***Offenheit (Big-Five, Satow, 2012)***

Um die Persönlichkeitseigenschaft *Offenheit* der Proband\*innen zu erheben, wurde der *Big-Five*-Persönlichkeitstest (B5T) von Satow (2012) verwendet.

Der B5T wurde entwickelt, um mit hoher Reliabilität die fünf grundlegenden Persönlichkeitsdimensionen – *Offenheit*, *Gewissenhaftigkeit*, *Extraversion*, *Verträglichkeit* und *Neurotizismus* - zu erfassen. Er ist kürzer als die meisten bisherigen Tests und deshalb für Online-Testungen, aufgrund geringer Drop-Out-Raten, gut geeignet (Bühner, 2010). Die für die Eichung getestete Stichprobe beträgt 5520 Personen.

Für die Untersuchung wurde nur der Teilbereich „Offenheit“ des Fragebogens herangezogen. Die Skala erreicht eine gute Reliabilität von Cronbach's Alpha .76 (Cronbach, 1951) und die faktorielle Validität ist ausreichend, 8 von 10 Items laden  $> .30$  auf den eigenen Faktor, nur 3 Items laden bei ungefähr  $.30$  auf den Faktor *Extraversion*. Die Items sind nachvollziehbar formuliert und für Proband\*innen ab 16 Jahren geeignet. Die Beantwortung der 10 Items erfolgte mittels vierstufiger Likert-Skalen, was den Vorteil hat, dass es den Teilnehmer\*innen nicht möglich war, eine „neutrale“, mittlere Antwortkategorie auszuwählen (Garland, 1991): 1 = „Trifft gar nicht zu“; 2 = „Trifft eher nicht zu“; 3 = „Trifft eher zu“; 4 = „Trifft genau zu“.

Folgend werden zwei Beispielitems zur Erfassung der *Offenheit* mittels B5T (Satow, 2012) dargestellt:

„Ich will immer neue Dinge ausprobieren.“

„Ich lese viel über wissenschaftliche Themen, neue Entdeckungen oder historische Begebenheiten.“

Sämtliche 10 Items sind mit den dazugehörigen wichtigsten Kennwerten im Kapitel „Anhang F: Items-*Offenheit* (B5T, Satow, 2012) in Tabelle 11 zu finden.

## ***Beurteilungsfehler bei der Leistungsbeurteilung (Halo-Effekt): Benotung eines Aufsatzes***

Um die Beurteilungsverzerrung zu testen, wurde den Lehrpersonen je Proband\*in eine von drei alternativen Vorgeschichten zu einem fiktiven 10-jährigen Schüler namens Peter als Instruktionen dargeboten. Bei Instruktionen A und B wurden für eine Benotung erbrachter Leistungen irrelevante Informationen zu Peter dargeboten, die die Benotung der Untersuchungsteilnehmer\*innen positiv oder negativ, nach dem Vorbild des *Halo-Effekts* (Thorndike, 1920; Cooper, 1981), beeinflussen sollten (Malouff & Thorsteinsson, 2016). Die Kontrollgruppe, Instruktion C, erhielt keine irrelevanten Informationen.

Anschließend wurden die Proband\*innen aufgefordert, einen Aufsatz des fiktiven Schülers mit den Ziffernnoten von 1 bis 5 zu benoten. Ein Texteingabefeld mit der Angabe „Falls Sie, für sich selbst, Notizen zur Benotung machen wollen, können Sie dafür dieses Eingabefeld nutzen:“, stand den Proband\*innen als Hilfestellung für die Leistungsbeurteilung zur Verfügung.

Weiters bestand anstatt der Benotung die Möglichkeit „kann ich nicht beurteilen“ auszuwählen. In diesem Fall konnten die Proband\*innen in einer freien Texteingabe formulieren, warum sie den Aufsatz nicht benoteten. Eine Zusammenfassung dieser qualitativ-explorativ erhobenen Daten wird in Kapitel 5.3. berichtet und im Kapitel 6.3 diskutiert.

Folgend werden die drei Vorgeschichten: Instruktionen A, B und C sowie der zu benotende Aufsatz dargestellt.

## Irrelevante Nebeninformationen

### Instruktion A (positiv):

Stellen Sie sich vor, Sie sind Klassenlehrer\*in einer 4. Klasse einer öffentlichen Volksschule. Sie sind gerade dabei, eine Aufgabe zu verbessern und zu benoten. Das Thema zur Geschichte, die die Kinder selbständig verfassen sollten, war "Erzähle ein Erlebnis, wo du Angst hattest". Sie verbessern gerade die Geschichte von Peter. Er ist 10 Jahre alt und kommt immer ordentlich gekleidet und ausgeschlafen in die Schule. Peter war seit Beginn der ersten Klasse einer Ihrer fleißigsten Schüler. Nie hat er eine Hausaufgabe vergessen zu erledigen und im Unterricht passte er auch immer auf.

### Instruktion B (negativ):

Stellen Sie sich vor, Sie sind Klassenlehrer\*in einer 4. Klasse einer öffentlichen Volksschule. Sie sind gerade dabei, eine Aufgabe zu verbessern und zu benoten. Das Thema zur Geschichte, die die Kinder selbständig verfassen sollten, war "Erzähle ein Erlebnis, wo du Angst hattest". Sie verbessern gerade die Geschichte von Peter. Er ist 10 Jahre alt und kommt meist unordentlich gekleidet und verschlafen in die Schule. Er war seit Beginn der ersten Klasse einer Ihrer faulsten Schüler. Er hat bisher viele Hausaufgaben vergessen zu erledigen und im Unterricht passte er auch sehr selten auf.

### Instruktion C (neutral/Kontrollgruppe):

Stellen Sie sich vor, Sie sind Klassenlehrer\*in einer 4. Klasse einer öffentlichen Volksschule. Sie sind gerade dabei, eine Aufgabe zu verbessern und zu benoten. Das Thema zur Geschichte, die die Kinder selbständig verfassen sollten, war "Erzähle ein Erlebnis, wo du Angst hattest". Sie verbessern gerade die Geschichte von Peter. Er ist 10 Jahre alt.

## Zu benotender Aufsatz: „Peters Geschichte“

"Der Tiger im Bett"

Eines Mittwochs des Dezembers schimpfte meine Mutter wieder einmal. Mein Bruder hatte noch Hausübung und ich wollte nicht ins Bett. Schließlich gab ich doch nach und ging ganz langsam hinauf. Als ich in mein Zimmer ging, sah ich etwas auf meinem Bett liegen. Wenn ich genauer hinschaute, sah es so aus als wäre es ein liegender Tiger. Ich wurde blass, totenblass. Im Zimmer war es nämlich stockfinster, und ich konnte nur wenig erkennen. Ich wollte mich ducken, doch da stieß ich an. Ich sah aus den Augenwinkeln, dass sich der Tiger bewegte, Und ich schrie "Hilfe". Das Herzklopfte mir bis zum Hals. Ich hatte angst, Todesangst und ich dachte dass er mich fressen würde. Ich rannte zum Lichtschalter und schaltete das Licht an. Als ich sah, dass der Tiger nicht anderes war als Decken und Polster, war Ich sehr erleichtert. Die Ohren waren die Sofapolster und der Körper die Bettdecke. Als meine Mutter endlich kam, Erzählte ich ihr alles und leider hatte ich seit dem Angst vor Tigern und anderenen Wildtieren.

### 4.3. Stichprobe

Die verwendete Stichprobe setzt sich aus 315 Personen zusammen, wobei der Datensatz um 23 Fälle bereinigt wurde, wovon 17 Fälle frühzeitig abgebrochen wurden und sechs Fälle doppelt vorhanden waren. In den Tabellen 3 und 4 werden die deskriptiven Charakteristiken der Stichprobe getrennt in Laien- sowie Lehrpersonen-Stichproben dargestellt:

Tabelle 3: Stichprobenzusammensetzung der Laien.

<b>n</b>	123
Altersspanne	18 - 78
Mittelwert Alter (SD)	44 (14.81)
<b>Geschlecht, n (%)</b>	
weiblich	83 (67.5)
männlich	40 (32.5)
<b>Bildungsstand, n (%)</b>	
Pflichtschulabschluss	1 (0.8)
mittlere berufsbildende Ausbildung	26 (21.1)
Hochschulreife	32 (26)
Hochschulabschluss	48 (39)

Tabelle 4: Stichprobenzusammensetzung der Lehrer\*innen.

<b>n</b>	192
Altersspanne	22 - 73
Mittelwert Alter (SD)	42 (12.54)
<b>Geschlecht, n (%)</b>	
weiblich	180 (93.8)
männlich	12 (6.3)
<b>Bildungsstand, n (%)</b>	
Hochschulabschluss (Pädagogische Akademie)	192 (100)
<b>Berufserfahrung, n (%)</b>	
Jahre	1 - 43
Mittelwert Jahre (SD)	16.39 (12.24)
Mehr Erfahrung mit Beurteilung durch Ziffernnoten	56 (29.2)
Mehr Erfahrung mit Verbalbeurteilungen	32 (16.7)
Ungefähr gleich viel Erfahrung mit Beurteilung durch Ziffernnoten und Verbalbeurteilungen	88 (45.8)
Deutsch als Unterrichtsfach	168 (87.5)
Bereits eine Weiterbildung im Bereich „Leistungsbeurteilung“ absolviert.	77 (40.1)

## 5. Ergebnisse

Die Hypothesen wurden wie folgt untersucht:

Hypothese 1: Chi<sup>2</sup>-Tests zur Untersuchung, ob es Unterschiede in der Beurteilungspräferenz – Verbal- oder Ziffernbeurteilung – zwischen Lehrer\*innen und Laien je Schulklasse gibt.

Hypothese 2: Korrelationsanalyse, ob es Zusammenhänge zwischen der Beurteilungspräferenz – Verbal- oder Ziffernbeurteilung – und der Ausprägung der *Offenheit* gibt.

Hypothese 3: Einfaktorielle Varianzanalyse (ANOVA), um zu untersuchen, ob es signifikante Unterschiede zwischen den drei Gruppen – positive, neutrale und negative Vorgeschichte - in der Benotung von Peters Geschichte durch die Lehrer\*innen gibt.

## **5.1. Unterschiede in den Beurteilungspräferenzen je Schulklasse von Lehrer\*innen und Laien**

Die zwei Gruppen „Lehrer\*innen“ und „Laien“ wurden mittels Chi<sup>2</sup>-Tests bezüglich der Verteilungen Ihrer Angaben, ob sie Verbalbeurteilungen oder Ziffernnoten für die vier Volksschulklassen bevorzugen, verglichen. Es wurden in sämtlichen Vergleichen der vier Schulklassen keine signifikanten Unterschiede zwischen diesen zwei Gruppen gefunden. Für die erste Klasse zeigen die Ergebnisse eine Bevorzugung von Verbalbeurteilung gegenüber Ziffernnoten (88.8 % der Lehrkräfte und 87.2 % der Laien). Im Vergleich der zweiten Klasse, nähern sich die Präferenzen einander an, es besteht aber der Trend hin zu Verbalbeurteilungen, wobei sich 68.4 % Lehrkräfte und 74.3 % Laien für eine Verbalbeurteilung entschieden. Ab der dritten Klasse kommt es zu einer Trendumkehr und es wurden, sowohl von Lehrkräften (64.3 %) sowie von Laien (56.6 %), Ziffernnoten als bevorzugt für die Leistungsbeurteilung der Schüler\*innen angegeben. Für die vierte Klasse entschieden sich 74.4 % der Lehrkräfte und 68.6 % der Laien für Ziffernnoten.

In den Tabellen 5, 6, 7 und 8 werden die nicht signifikanten Ergebnisse der Untersuchungen, ob es Unterschiede in der Beurteilungspräferenz zwischen Lehrer\*innen und Laien je Schulklasse gibt, dargestellt.

Tabelle 5: Präferenz Leistungsbeurteilung 1.Klasse:

	Verbalbeurteilung	Ziffernnoten	Gesamt
Lehrer*innen	135 (88.8 %)	17 (11.2 %)	152
Laien	95 (87.2 %)	14 (12.8 %)	109
Gesamt	230 (88.1 %)	31 (11.9 %)	261

Tabelle 5: Chi<sup>2</sup>-Test: Darstellung der Verteilung der präferierten Leistungsbeurteilung (Verbalbeurteilung oder Ziffernnoten) bei Lehrer\*innen und Laien: n = 261, p = .68, Cramer-V = .03

Tabelle 6: Präferenz Leistungsbeurteilung 2.Klasse:

	Verbalbeurteilung	Ziffernnoten	Gesamt
Lehrer*innen	104 (68.4 %)	48 (31.6 %)	152
Laien	81 (74.3 %)	28 (25.7 %)	109
Gesamt	185 (70.9 %)	76 (29.1 %)	261

Tabelle 6: Chi<sup>2</sup>-Test: Darstellung der Verteilung der präferierten Leistungsbeurteilung (Verbalbeurteilung oder Ziffernnoten) bei Lehrer\*innen und Laien: n = 261, p = .30, Cramer-V = .06

Tabelle 7: Präferenz Leistungsbeurteilung 3.Klasse:

	Verbalbeurteilung	Ziffernnoten	Gesamt
Lehrer*innen	55 (35.7 %)	99 (64.3 %)	154
Laien	46 (43.4 %)	60 (56.6 %)	106
Gesamt	101 (38.8 %)	159 (61.2 %)	260

Tabelle 7: Chi<sup>2</sup>-Test: Darstellung der Verteilung der präferierten Leistungsbeurteilung (Verbalbeurteilung oder Ziffernnoten) bei Lehrer\*innen und Laien: n = 260, p = .21, Cramer-V = .08

Tabelle 8: Präferenz Leistungsbeurteilung 4.Klasse:

	Verbalbeurteilung	Ziffernnoten	Gesamt
Lehrer*innen	39 (25.3 %)	115 (74.4 %)	154
Laien	33 (31.4 %)	72 (68.6 %)	105
Gesamt	72 (27.8 %)	187 (72.2 %)	259

Tabelle 8: Chi<sup>2</sup>-Test: Darstellung der Verteilung der präferierten Leistungsbeurteilung (Verbalbeurteilung oder Ziffernnoten) bei Lehrer\*innen und Laien: n = 259, p = .28, Cramer-V = .07

n = 46 (30%) der Lehrer\*innen und n = 30 (27.5%) der Laien gaben zusätzlich zur Angabe der Beurteilungspräferenz eine schriftliche Begründung Ihrer Auswahl an.

Die Aussagen der Lehrer\*innen beziehen sich zu einem großen Teil auf die Vorteile von Verbalbeurteilungen. Es werden die Möglichkeiten individueller und differenzierter Beurteilungen durch Verbalbeurteilungen hervorgehoben. Verbalbeurteilungen seien im Gegensatz zu Noten im Stande, auch das Bemühen, die soziale Kompetenz und den aktuellen Leistungsstand der Schüler\*innen detailliert zu beschreiben.

Die Laien hoben ebenso die vielseitigen Möglichkeiten der Leistungsrückmeldung durch Verbalbeurteilungen als positives Kriterium hervor. Kinder sollten so lange wie möglich vom gesellschaftlichen Leistungsdruck abgeschirmt werden.

Die vollständigen Auflistungen der Begründungen sind unter „Anhang G: Begründungen für die Beurteilungspräferenzen (Lehrer\*innen)“ sowie „Anhang H: Begründungen für die Beurteilungspräferenzen (Laien)“ in den Tabellen 12 und 13 zu finden.

## 5.2. Der Zusammenhang zwischen *Offenheit* und Beurteilungspräferenz

Es wurden punktbiseriale Korrelationsanalysen durchgeführt, um zu untersuchen, ob die Variablen „Offenheit“ und „Beurteilungspräferenz“ je Schulklasse einen signifikanten Zusammenhang aufweisen. Für die Untersuchung wurde der Mittelwert des *Offenheit*-Scores (Satow, 2012) und die Beurteilungspräferenz herangezogen.

Die Variable *Offenheit* (N = 273) wurde mittels Kolmogorov-Smirnov Test negativ auf Normalverteilung getestet:  $p < .001^{***}$  (M = 3.05, SD = .43).

Die „Beurteilungspräferenz“ wurde in die Kategorien „Verbalbeurteilungen“ und „Ziffernnoten“ eingeteilt. Eine Veranschaulichung der Ergebnisse ist in Tabelle 9 zu finden.

Tabelle 9: Veranschaulichung der Ergebnisse der bivariaten Zusammenhänge:

	1. Klasse Verbalbeurteilung/ Ziffernnoten	2. Klasse Verbalbeurteilung/ Ziffernnoten	3. Klasse Verbalbeurteilung/ Ziffernnoten	4. Klasse Verbalbeurteilung/ Ziffernnoten
<i>Offenheit</i>	.11* (p .045)	.15* (p .017)	.23*** (p < .001)	.12* (p .048)

Tabelle 9: Korrelationen der Variablen „Offenheit“ und „Beurteilungspräferenz“ je Schulklasse. Alle vier Untersuchungen weisen signifikante positive Korrelationen zwischen *Offenheit* und der Präferenz für Verbalbeurteilungen auf. Anmerkung: \* =  $p < .05$ ; \*\* =  $p < .01$ ; \*\*\* =  $p < .001$

### 5.3. Beurteilungsfehler bei der Benotung (*Halo-Effekt*)

Die Mittelwerte der Benotungen der „Lehrer\*innenstichprobe“ (n = 173) der Gruppen „positive Instruktion“, „neutrale Instruktion“ und „negative Instruktion“ wurden miteinander verglichen, um zu untersuchen, ob es signifikante Unterschiede in den drei Gruppen gibt. Die berechnete einfaktorielle ANOVA ( $F = .08$ ,  $p = .93$ ) lässt auf keine systematischen Unterschiede zwischen den drei Gruppen schließen. Eine Auflistung der deskriptiven Kennwerte je Gruppe ist in Tabelle 10 dargestellt. Die Verteilung der gesamten Benotungen von Peters Geschichte ist in Abbildung 1 zu sehen.

Tabelle 10: Deskriptive Statistiken der Benotung von Peters Geschichte je Gruppe:

	Mittelwert	SD (SF)	Minimum – Maximum
Positive Instruktionen	2.32	.71 ( .10)	1 - 4
Neutrale Instruktionen	2.32	.64 ( .08)	1 - 4
Negative Instruktionen	2.30	.65 ( .09)	1 - 3
Gesamt	2.30	.66 ( .05)	1 - 4

Tabelle 10: Darstellung von Mittelwert, Standardabweichung (SD) und Standardfehler (SF) für die Gruppen „positive Instruktionen“, „negative Instruktionen“ und „neutrale Instruktionen“.

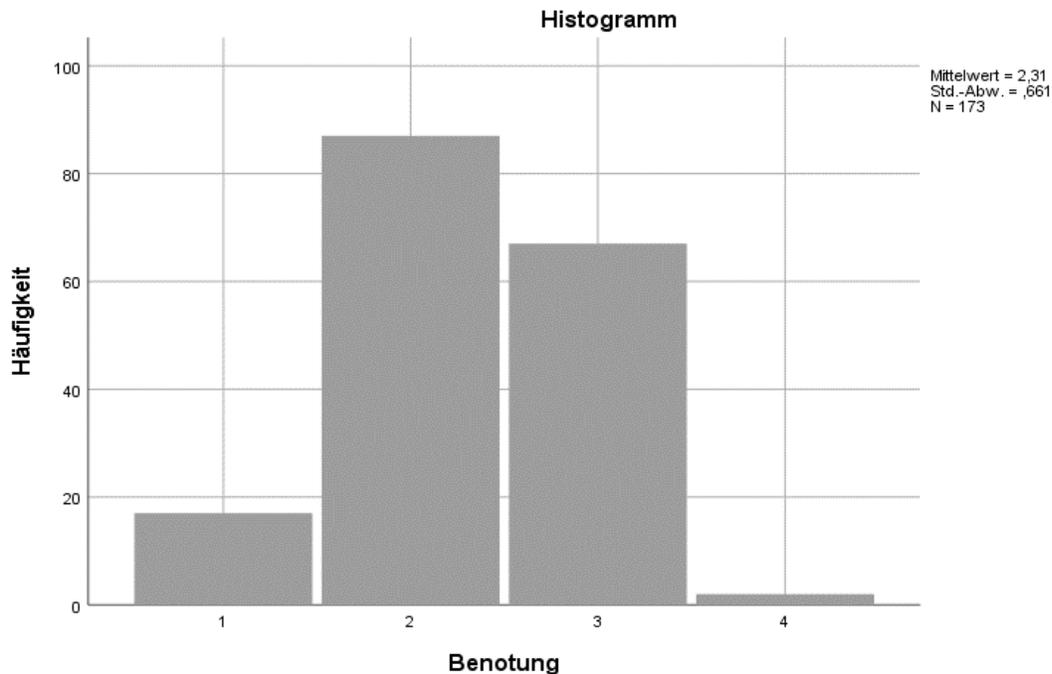


Abbildung 1: 148 (77.1 %) der Lehrer\*innen benoteten Peters Geschichte: Note 1 (14, 7.3 %), Note 2 (74, 38.5 %), Note 3 (58, 30.2 %), Note 4 (2, 1 %), Note 5 (0)

N = 44 (22.9 %) benoteten Peters Geschichte nicht:

Ohne Angabe einer Begründung nicht benotet: 34 (17.7 %),

mit Angabe einer Begründung nicht benotet: 10 (5.2 %).

Zehn Lehrer\*innen benoteten die Geschichte nicht und gaben unter anderem als Begründung dafür die fehlende soziale Bezugsnorm, die fehlende Kenntnis über die vorangehenden Leistungen von Peter oder das Fehlen einer sachlichen Bezugsnorm an.

Die 10 Begründungen, für die nicht benoteten Geschichten sind im Anhang unter „Anhang I: Begründungen für die Nicht-Benotung von Peters Geschichte“ zu finden.

## 6. Diskussion

Mithilfe der Studie wurde 1.) untersucht, ob sich die Leistungsbeurteilungspräferenz – Verbalbeurteilungen oder Beurteilung durch Ziffernnoten - für die vier Klassen der Volksschule zwischen Lehrer\*innen und Laien unterscheiden. 2.) Wurde versucht, mutmaßliche signifikante Zusammenhänge zwischen den Variablen „*Offenheit*“ und „Beurteilungspräferenz“ im Hinblick auf die einzelnen Schulklassen der Volksschule zu ermitteln. Es wurde angenommen, dass eine starke Ausprägung des Merkmals *Offenheit* mit einer Verbalbeurteilungspräferenz positiv korreliert. 3.) Wurde untersucht, ob Lehrkräfte sich bei Ihrer Benotung von Peters Geschichte durch vorangehende für die Beurteilung irrelevante Informationen über Peter beeinflussen lassen.

### 6.1. Unterschiede in den Beurteilungspräferenzen je Schulklasse von Lehrer\*innen und Laien

Die sehr ähnlichen Verteilungen der Angaben für die Beurteilungspräferenzen zwischen Laien und Lehrkräften sprechen für einen grundsätzlichen Konsens zwischen den zwei Gruppen. Das Ergebnis in der ersten Klasse zeigt eine klare Favorisierung der Verbalbeurteilung für Schuleinsteiger\*innen. Das Ergebnis deckt sich auch mit den aus der Literatur bekannten Annahmen (Lütgert et al., 2001; Valtin et al., 2002), dass Schulnoten nicht für die heterogenen Zusammensetzungen der Kinder und Förderbedarfe in der Volksschule ausreichend sind. Die Verbalbeurteilungen ermöglichen in den ersten zwei Klassen eine differenzierte Sichtweise (Black & William, 2010) der Fähigkeiten der Kinder und unterstützen sie betreffend die Leistungsmotivation und können somit einen erfolgreichen Einstieg in die individuellen Bildungskarrieren der Kinder ermöglichen. Das Ergebnis der dritten Klasse zeigt ein nicht mehr so eindeutiges Bild. Lehrer\*innen stehen mit ihrem Verhältnis (64.3 % Ziffernnoten zu 35.7 % Verbalbeurteilungen) im Gegensatz zu den restlichen Proband\*innen (56.6 % Ziffernnoten zu 43.4 % Verbalbeurteilungen) mit einem eindeutigeren Ergebnis da. Möglicherweise kommt hier ein Effekt durch die

Wiedereinführung der Ziffernnoten ab dem Jahreszeugnis der zweiten Volksschulklasse zum Tragen. Laien könnte diese Wiedereinführung eventuell gar nicht bekannt sein und daher ein unbeeinflusster Zugang zur neuerlichen Reformation zuteilwerden. Der Trend hin zu Ziffernnoten erreicht in der vierten Schulklasse seinen Höhepunkt und kommt sowohl bei Lehrer\*innen als auch bei den restlichen Proband\*innen zu einem eindeutigen Ergebnis. Hier könnte vor allem die Allokationsfunktion von Ziffernnoten (Paasch et al., 2019) ausschlaggebend sein. Die Vorgabe, weiterführende Schulen wie das Gymnasium, nur aufgrund eines guten Notendurchschnittes absolvieren zu dürfen, könnte mitunter eine Begründung für dieses Ergebnis sein. Allein die Tatsache, dem Bildungssystem und seinen Beurteilungsmethoden Folge leisten zu müssen, könnte ein Grund dafür sein, sich nicht für die differenziertere Verbalbeurteilung zu entscheiden, sondern die inhaltsarme (BIFIE, 2019; Harwell & Gatti, 2001) und fehleranfällige Beurteilungsmethode (Malouff & Thorsteinsson, 2016) der rangskalierten Noten zu wählen.

### **Begründungen der Leistungsbeurteilungspräferenzen**

Die Begründungen der Leistungsbeurteilungspräferenzen lassen sich wie folgt kurz zusammenfassen: Die Aussagen der Lehrer\*innen beziehen sich in erster Linie auf die Vorteile von Verbalbeurteilungen. Es werden die Möglichkeiten differenzierte Beurteilungen der Stärken, Schwächen und des Aufholbedarfes der Schüler\*innen hervorgehoben. Verbalbeurteilungen seien im Gegensatz zu Noten im Stande, auch das Bemühen, die soziale Kompetenz und den aktuellen Leistungsstand der Schüler\*innen detailliert zu beschreiben. Andererseits präferierten viele Proband\*innen, vor allem ab dritter oder vierter Klasse die Ziffernnoten. Das aktuelle Schulsystem mit den an Ziffernnoten gekoppelten Selektionsverfahren und der weiteren omnipräsenten Anwendung von Ziffernnoten in sämtlichen Schulen lasse keine alternativen Beurteilungsmethoden zu. Hinzu komme, dass Ziffernnoten von manchen Eltern aufgrund mangelnder Sprachkenntnisse oder anderer Gründe bevorzugt werden oder dass Verbalbeurteilungen ohnehin nur Individualität vortäuschten, aber in Wirklichkeit nur nichtssagende Phrasen seien.

Folgend werden die Aussagen der Laien zusammengefasst: Es wurden vorwiegend die Möglichkeiten der differenzierten Leistungsrückmeldung von Verbalbeurteilungen als positives Kriterium hervorgehoben. Kinder sollten so lange wie möglich vom gesellschaftlichen Leistungsdruck abgeschirmt werden. Ziffernnoten förderten den Vergleichswettkampf und Verbalbeurteilungen werden von Kindern besser verstanden. Wenige Untersuchungsteilnehmer\*innen sprachen sich für die Notwendigkeit von Ziffernnoten aus, um die Schüler\*innen auf den weiteren Bildungsweg vorzubereiten, oder dass beide Beurteilungsmethoden angewandt werden sollten.

Sowohl Lehrer\*innen als auch Laien sind sich bei der Präferenz für Verbalbeurteilungen wegen der Möglichkeit differenzierter Leistungsrückmeldungen und der aufgrund des Schulsystems bzw. des gesellschaftlichen Leistungsdrucks bevorzugten Ziffernnoten in den vierten Schulklassen einig. Lehrer\*innen merkten mehrfach an, dass die Sprachkompetenz der Eltern der zu beurteilenden Kinder eine wesentliche Rolle spiele und somit die Ziffernbeurteilung wiederum an Präferenz gewann. Grundsätzlich gibt es aber, wie schon bereits weiter oben angemerkt, einen Konsens in der verbalen Leistungsbeurteilungspräferenz der ersten zwei Klassen. Die Leistungsbeurteilungspräferenz für die dritte Klasse - der Proband\*innen, die eine Erklärung für ihre Präferenz verschriftlichten - weist im Gegensatz zur Gesamtstichprobe ebenso eine verbale Leistungsbeurteilungspräferenz auf. Dies gilt sowohl für Lehrer\*innen und sonstige Untersuchungsteilnehmer\*innen.

## 6.2. Der Zusammenhang zwischen *Offenheit* und Leistungsbeurteilungspräferenz

Die *Offenheit* wurde mit einem Kurzinventar mit 10 Fragen erhoben (Satow, 2012). Es wurde versucht, mithilfe von punktbiserialen Korrelationsanalysen Zusammenhänge zwischen den Variablen „*Offenheit*“ und „Beurteilungspräferenz“ (Verbalbeurteilung oder Ziffernnoten) je Schulklasse zu finden. Es wurden sowohl Lehrpersonen als auch Laien für diese Untersuchung herangezogen, weil Eltern von Volksschüler\*innen mitentscheiden können, ob ihre Kinder bis zum Ende des ersten Semesters der zweiten Volksschulklasse Verbal- oder Ziffernbeurteilungen erhalten. Es zeigten sich bei allen vier Schulklassen signifikante Ergebnisse, die für einen positiven Zusammenhang zwischen Präferenz von Verbalbeurteilungen und *Offenheit* sprechen: Erste Klasse: .11\*; Zweite Klasse: .15\*; Dritte Klasse .23\*\*\*; Vierte Klasse: .12\*. In der Literatur wird eine hohe Ausprägung von *Offenheit* mit Kreativität (Tan et al., 2019; Erez & Nouri, 2010), Interesse für komplexe und nicht-lineare Aufgaben (Faust & Foerster, 2008) sowie *Ambiguitätstoleranz* (Hartinger et al., 2005) in Zusammenhang gebracht.

In der ersten ( $r = .11^*$ ) und zweiten Klasse ( $r = .15^*$ ) zeigt die Untersuchung positive Zusammenhänge in einem geringen Ausmaß zwischen der Ausprägung von *Offenheit* und der Verbalbeurteilungspräferenz. In der dritten Klasse ( $r = .23^{***}$ ) zeigt sich im Gegensatz zu den vorangehenden Klassen ein leicht erhöhter hoch signifikanter Zusammenhang. Diese Tatsache könnte ein Hinweis darauf sein, dass sich Personen mit ausgeprägter *Offenheit* eher über die Sinnhaftigkeit von Leistungsbeurteilungen Gedanken machen, ohne sie für gegeben hinzunehmen. In der vierten Klasse ( $r = .12^*$ ) verringert sich der signifikante Zusammenhang zwischen *Offenheit* und Verbalbeurteilungspräferenz wiederum. Möglicherweise nimmt dieser Wert ein geringeres Ausmaß an, weil die Perspektive für eine weitgreifende Lösung im Sinne der Verbalbeurteilungsmodalität in unserem Schulsystem fehlt. Die Notwendigkeit (gute) Noten zu erhalten, um im Bildungssystem weiter zu kommen, ist omnipräsent. Somit bestehen wahrscheinlich auch nur geringe Chancen, dass in den Volksschulen

(spätestens in den vierten Klassen) auf Persönlichkeitsentwicklung und elaborierenden Lernmethoden abzielende Leistungsbeurteilungsmethoden in Betracht gezogen werden können.

### **6.3. Beurteilungsfehler bei der Benotung (*Halo-Effekt*)**

Der sogenannte *Halo-Effekt* konnte in dieser Untersuchung nicht repliziert werden (Malouff & Thorsteinsson, 2016). Es wurden keine signifikanten Zusammenhänge zwischen den drei Vorgeschichten (Instruktionen A, B, C) zu Peter und der jeweiligen Benotung seiner erbrachten Leistung bzw. seiner verfassten Geschichte gefunden. Mit den Vorgeschichten wurde versucht, die Beurteilung von Peters Geschichte durch die Lehrer\*innen mit jeweils entweder einem positiven, einem negativen oder gar keinem *Halo* bzw. irrelevanten Informationen zu verknüpfen, um die Benotung zu beeinflussen. Die berechnete ANOVA zeigte keinen Nachweis für einen Zusammenhang zwischen den Vorgeschichten und den Noten ( $F = .08$ ,  $p .93$ ). Was aber trotz des nicht signifikanten Ergebnisses den *Halo-Effekt* betreffend auffällt, ist, dass es eine hohe Variabilität in der Verteilung der Ziffernnoten gibt. Von den 173 (77.1 %) der Lehrpersonen, die Peters Geschichte benoteten, fielen die Noten wie folgt aus: 1 (14, 7.3 %); 2 (74, 38.5 %); 3 (58, 30.2 %); 4 (2, 1 %); 5 (0). Dieses Ergebnis unterstützt die bereits aus der Literatur bekannte Annahme der unzureichenden Messgenauigkeit von Ziffernnoten (BIFIE, 2019).

Von den Lehrer\*innen, die keine Beurteilung aber eine Begründung für die fehlende Leistungsbeurteilung abgaben, war auffällig, dass sich Lehrkräfte im beruflichen Alltag an unterschiedlichen Rahmenbedingungen zur Benotung orientierten. Es werden bspw. Referenzrahmen, wie die soziale und individuelle Bezugsnorm (Häcker et al., 2013), Kriterienkataloge und Beurteilungstabellen herangezogen. Auf Basis dieser Einzelleistungen werden aggregierte Leistungen der Schüler\*innen, die kaum noch Nachvollziehbarkeit erweisen, für den Aufstieg in die nächste Schule verwendet.

## 7. Limitationen

Die Stichprobe ist selbstselektiert, es wurden via Online-Plattformen und E-Mails Einladungen für die Studie versandt. Es gibt keine eindeutig nachvollziehbaren Kriterien für die Teilnahme der Proband\*innen an der Untersuchung. Lehrer\*innen anderer Schulformen (Sekundarschule, etc.) könnten die Ergebnisse aufgrund der Vergleiche zwischen Volksschullehrer\*innen und Laien verzerrt haben.

Bei den an der Untersuchung teilgenommenen Proband\*innen handelt es sich um eine homogene Stichprobe. Von den 315 Personen sind 263 (83.5 %) weiblich und 52 (16.5 %) männlich. Die Stichprobe setzt sich hauptsächlich aus Personen mit höheren Bildungsabschlüssen zusammen. Personen mit geringerem Bildungsgrad und männliche Probanden sind in der Stichprobe unterrepräsentiert. Ungefähr 2/3 der Proband\*innen waren als Lehrpersonen an Volksschulen beschäftigt (213, 67.7 %). Diese einseitige Verteilung könnte für eine Verzerrung der Ergebnisse gesorgt haben.

Möglicherweise ergab sich ein weiteres Problem durch das Fragebogendesign des B5T zur Erfassung der *Offenheit* (Satow, 2012). 9 von 10 Items des Fragebogens wurden von den Proband\*innen nach demselben Schema durchgeführt. Es wurden 4-stufige Likert-Skalen von 1 = geringste Ausprägung und aufsteigend verwendet. Im Gegensatz dazu wurde die negative Polung von Item Nummer 4 „Am liebsten ist es mir, wenn alles so bleibt, wie es ist.“ beibehalten. Möglicherweise wurde die Beantwortung dieses Items, trotz der sehr kurzen Testdauer, von den Teilnehmer\*innen unterschiedlich durchgeführt.

Eine grundlegende Verfälschung von Persönlichkeitstests, wie auch die Erfassung der Persönlichkeitseigenschaft *Offenheit*, unterliegt einer tendenziell positiveren Selbstdarstellung aufgrund sozial erwünschtem Antwortverhalten. Versuche, sich weitestgehend positiv darzustellen, beeinträchtigen die Aussagekraft der erfassten Datensätze (Satow, 2012).

In dieser Studie wurde mithilfe von Nebeninformationen versucht eine mögliche Beurteilungsverzerrung bei der Benotung eines Aufsatzes zu untersuchen.

Möglicherweise ist das nicht signifikante Ergebnis darauf zurückzuführen, dass willkürlich erstellte Informationen und kein validiertes Werkzeug zur Erfassung des *Halo-Effektes* verwendet wurden. Weiters benoteten die Lehrer\*innen nicht einen Aufsatz einer/eines eigenen Schülers/Schülerin, sondern den einer fiktiven Person. Zusätzlich besteht die Möglichkeit, dass der Versuch der Beeinflussung der Benotungen durch die irrelevanten Nebeninformationen von den Lehrer\*innen erkannt wurde. Somit könnte der Einfluss eines eventuellen *Halo-Effektes* abgeschwächt worden sein. Eine Möglichkeit bei weiteren Untersuchungen wäre es, nach dem Abschluss des jeweiligen Fragebogens den Untersuchungsteilnehmer\*innen die Frage zu stellen, was sie für die Hintergründe der einzelnen Fragestellungen halten. Diese Informationen könnten sich als nützlich für die Auswertung der Ergebnisse herausstellen, egal ob es signifikante Ergebnisse gibt oder nicht (Malouff & Thorsteinsson, 2016).

## 8. Fazit

Die Studie kommt zu dem Ergebnis, dass es signifikante Zusammenhänge zwischen *Offenheit* und der Beurteilungspräferenz für Verbalbeurteilungen gegenüber Ziffernnoten in sämtlichen Klassen der Volksschule gibt. Die Korrelationen weisen je Schulklasse kleine bis mittlere Effektstärken auf (Cohen, 1992).

Es wurden keine statistisch signifikanten Unterschiede zwischen Volksschullehrer\*innen und Laien im Hinblick auf die Leistungsbeurteilungspräferenzen der einzelnen Schulklassen festgestellt. Eine explorative Datenanalyse zeigt, dass Proband\*innen, die eine Begründung für ihre Entscheidungen angaben, auch noch in der dritten Klasse vorwiegend die verbale Leistungsbeurteilung präferierten.

Der *Halo-Effekt* konnte in dieser Untersuchung nicht nachgewiesen werden. Es bestehen keine Zusammenhänge zwischen der für die Benotung von Peters Geschichte im Vorhinein dargebotenen irrelevanten Informationen und der gegebenen Note durch die Lehrer\*innen. Was aber durch das Ergebnis deutlich wird, ist die große Variabilität der vergebenen Notenbeurteilungen von 1 bis 4. Dies weist auf die in der Literatur bereits bekannten schlechten Gütekriterien von Ziffernnoten hin (BIFIE, 2019) und spricht ohne Zweifel für eine große Problematik, was die Allokationsfunktion von Ziffernnoten für die weiterführenden Schulen betrifft (Paasch et al., 2019).

Aus den gesammelten Daten geht ein gesellschaftlicher Konsens in der Präferenz der verbalen Beurteilung von Volksschüler\*innen hervor. Sowohl die statistischen Analysen als auch die explorativen Ergebnisse zeigen ein einheitliches Bild bei Lehrer\*innen und Laien. Eine ausgeprägte *Offenheit* (Satow, 2012) lässt sich mit dem Entschluss für eine individuellere und differenziertere Leistungsbeurteilungsmethode in Verbindung bringen. Die relativ kleinen Effekte sprechen dafür, dass diese Präferenzen weitgreifenden Erklärungen bedürfen. Unter anderem gibt es im bestehenden österreichischen Schulsystem, vor allem aufgrund der kurzen Grundschulzeit und der auf dem Notendurchschnitt basierenden Aufstiegsmöglichkeiten in weiterführende Schulen, kaum realistische Lösungsansätze für alternative Beurteilungsmethoden ab der vierten

Volksschulklasse. Die Bildungssysteme vieler laut PISA (OECD, 2019) erfolgreichen Länder nutzen Schulmodelle, die gesamtschulartig aufgebaut sind und über lange Zeiträume auf Ziffernnoten verzichten. Die Leistungsbeurteilungen der Lehrer\*innen dienen fast ausschließlich der Förderung individueller Zielsetzungen und auf selektive Beurteilungen wird verzichtet (Internationale Vergleichsstudie, 2003).

Gegen eine parallele Anwendung von Verbalbeurteilungen und Ziffernnoten spricht der von Black und Wiliam (2010) beschriebene Effekt, dass die Aufmerksamkeit der Schüler\*innen von den inhaltsvollen verbalen Beurteilungen weg auf die leichter vergleichbareren Ziffernnoten gezogen wird. Nach der Erwartungs-Mal-Wert-Theorie von Eccles (2005) würden Schüler\*innen sich fast ausschließlich darauf konzentrieren, bessere Noten zu bekommen, ohne dabei das eigentlich zu Lernende vertiefen zu wollen. Somit würden die zuständigen Lehrer\*innen einen doppelten Aufwand für kaum bis gar keinen Mehrwert leisten müssen und die Gefahr für stereotype Äußerungen bei der Beurteilung würde steigen (Schmude, 2005; Valtin, 2002).

Bildungsziele werden umfangreicher und können mit den bestehenden schulfachspezifischen Ziffernnoten nicht zufriedenstellend erfasst werden. Trotzdem wurde nicht versucht, diese Problemlage in Österreich in den jüngeren Schulreformen zu lösen, vielmehr wurde dies weiterhin ignoriert (Spiel & Wagner, 2003) und daher besteht, bereits in den Volksschulen beginnend, eine Überforderung unseres Schulsystems (Heid, 1999).

Das Ergebnis dieser Studie unterstützt die bereits aus der Literatur bekannten empirischen Nachweise die entgegengesetzt zu der in Österreich erneut eingeführten verpflichtenden Beurteilung durch Ziffernnoten ab dem Jahreszeugnis der zweiten Volksschulklassen stehen. Wie bereits erwähnt, sprechen viele empirische Nachweise dafür, eine verbale Beurteilung der erbrachten Leistung von Grundschüler\*innen zu forcieren, um optimale Voraussetzungen für ihre heterogenen Entwicklungsverläufe in Betracht der Bildungskarriere zu ermöglichen (BIFIE, 2019; Helmke, 2007; Yerkes & Dodson, 1908).

## 9. Abbildungsverzeichnis

Abbildung 1: Verteilung der Benotungen (1 bis 5) von Peters Geschichte durch Lehrpersonen.....	64
--	----

## 10. Tabellenverzeichnis

Tabelle 1: §14 der LBVO .....	19
Tabelle 2: Fünf Schlüsselstrategien der formativen Leistungsbeurteilung.....	26
Tabelle 3: Stichprobenzusammensetzung der Laien .....	55
Tabelle 4: Stichprobenzusammensetzung der Lehrer*innen .....	56
Tabelle 5: Präferenz Leistungsbeurteilung 1.Klasse .....	59
Tabelle 6: Präferenz Leistungsbeurteilung 2.Klasse.....	59
Tabelle 7: Präferenz Leistungsbeurteilung 3.Klasse.....	60
Tabelle 8: Präferenz Leistungsbeurteilung 4.Klasse.....	60
Tabelle 9: Veranschaulichung der Ergebnisse der bivariaten Zusammenhänge.....	62
Tabelle 10: Deskriptive Statistiken der Benotung von Peters Geschichte je Gruppe.....	63
Tabelle 11: Veranschaulichung der Items zur Kategorie „Offenheit“.....	99
Tabelle 12: Begründungen der Präferenzen der Lehrer*innen (n = 46).....	100
Tabelle 13: Begründungen der Präferenzen der Laien (n = 30).....	104

## 11. Literaturverzeichnis

Barrick, M. R., Mount, M. K., Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *Personality and Performance*, 9, S. 9–30.

Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K. J., Weiß, M. (2001): PISA. Programme for International Student Assessment (2001).

BIFIE, Bundesinstitut für Bildungsforschung, Innovation und Entwicklung des Bildungswesens, (2019). *Nationaler Bildungsbericht Österreich 2018, Band 1, Das Schulsystem im Spiegel von Daten und Indikatoren*. Graz: Leykam, DOI: <http://doi.org/10.17888/nbb2018-1.3>

Black, P. & William, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81–90. <https://doi.org/10.1177/003172171009200119>

Butler, R. (2005). Ein zielorientiertes Modell für Feedbackverfahren im Unterricht. Implikationen für Lernmotivation und Schulstruktur. *Unterrichtswissenschaft*, 33(2), S. 122-142.

Bühner, M. (2010). Einführung in die Test- und Fragebogenkonstruktion. Pearson Studium.

Bürgermeister, A., & Saalbach, H. (2018). Theoretischer Beitrag: Formatives Assessment: Ein Ansatz zur Förderung individueller Lernprozesse. *Psychologie in Erziehung und Unterricht*, 65(3), 194-205.

Brennan, D. J. (2008). University student anonymity in the summative assessment of written work. *Higher Education Research & Development*, 27(1), 43-54.

- Camadan, F., Reisoglu, I., Ursavas, Ö. F., & Mcilroy, D. (2018). How teachers' personality affect on their behavioral intention to use tablet PC. *International Journal of Information and Learning Technology*, 35(1), 12–28. <https://doi.org/10.1108/IJILT-06-2017-0055>
- Carless, D., & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, 43(8), S. 1315-1325.
- Cattell, R. B. (1957). *Personality and Motivation Structure and Measurement*. New York: World Book
- Cattell, R. B. (1965). *The Scientific Analysis of Personality*. London: Penguin.
- Cohen, J. (1992). A power primer. *Psychological bulletin*, 112(1), 155.
- Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin*, 90, 218–244. doi: 10.1037/0033-2909.90.2.218
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests\*. *Psychometrika*, 16(3).
- Deutsche Kultusministerkonferenz (2007). *Das Bildungswesen in der Bundesrepublik Deutschland 2006*. Bonn: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland. Zugriff am 02.02.2020. Verfügbar unter [http://www.kmk.org/fileadmin/doc/Dokumentation/Bildungswesen\\_pdfs/primarbereich.pdf](http://www.kmk.org/fileadmin/doc/Dokumentation/Bildungswesen_pdfs/primarbereich.pdf)
- Dresel, M. (2010). Förderung der Lernmotivation mit attributionalem Feedback. In C. Spiel, B. Schober, P. Wagner & R. Reimann (Hrsg.), *Bildungspsychologie* (S. 131-135). Göttingen: Hogrefe.
- Eder, F. (2008). Persönlichkeitsmerkmale von Lehramtsstudierenden. In *Gerechtigkeit und Effizienz im Bildungswesen. Unterricht, Schulentwicklung und LehrerInnenbildung als professionelle Handlungsfelder* (S. 273-293). LitVerlag.
- Eder, F., Neuweg, G. H., & Thonhauser, J. (2009). Leistungsfeststellung und Leistungsbeurteilung. *Bildungsbericht 2009*, (371), S. 247–268.

Eulenberger, J. (2015). Die Persönlichkeitsmerkmale von Personen im Kontext des Lehrer\_innenberufs (Personality Traits of Teaching Professionals).

Erez, M. & Nouri, R. (2010). Creativity: The influence of cultural, social, and work contexts. *Management and Organization Review*, 6, S. 351–370; doi:10.1111/j.1740-8784.2010.00191

Eysenck, H. J. (1967). *Biological basis of personality*. Charles C Thomas Publisher, Springfield, Illinois, USA.

Eysenck, H. J. (1947). *Dimensions of Personality*. London: Routledge & Kegan Paul.

Faust, G., & Foerster, F. (2008). Offenheit für Ideen und Handlungen bei Grundschullehramtsstudierenden. *Veränderte Kindheit: Konsequenzen für die Lehrerbildung*, S. 129-139.

Fayn, K., Silvia, P. J., MacCann, C., & Tiliopoulos, N. (2017). Interested in different things or in different ways? Exploring the engagement distinction between openness and intellect. *Journal of Individual Differences*, 38(4), S. 265–273. <https://doi.org/10.1027/1614-0001/a000243>

Feeley, T. H. (2002). Comment on halo effects in rating and evaluation research. *Human Communication Research*, 28(4), S. 578-586.

Fend, H. (2008). *Neue Theorie der Schule* (2. Aufl.). Wiesbaden: VS Verlag.

Figlio, D. N. (2005). *Names, expectations and the black-white test score gap* (No. w11195). National Bureau of Economic Research. Filippin, A., & Paccagnella, M. (2012). Family background, self-confidence and economic outcomes. *Economics of Education Review*, 31(5), S. 824-834.

Fischbach, A., Brunner, M., Krauss, S., & Baumert, J. (2016). Die Bezugsnormorientierung von Mathematiklehrkräften am Ende der Sekundarstufe I: Konvergenz verschiedener Messverfahren und Wirkung auf motivational-affektive Aspekte des Mathematiklernens und Leistung. *Journal for Educational Research Online*, 7(3), S. 3-27.

Freese, H. L. (1990): Verbale Beurteilung auf Jahrgangsstufe 3. Abschlussbericht über eine Begleituntersuchung. Berlin. Fuchs, L. S., Fuchs, D., Karns, K., Hamlett, C. L.,

Garland, R. (1991) The Mid-Point on a Rating Scale: Is it Desirable? *Marketing Bulletin*, 1991, 2, S. 66-70, Research Note 3.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In Wheeler (Hrsg.), *Review of Personality and social psychology*, Vol. 1, S. 141–165. Beverly Hills, CA.

Götz, T., Hall, N. C., Martiny, S. E., & Keller, M. (2013). Emotion and achievement in the classroom. *International Guide to Student Achievement*, 251, S. 192-195.

Hattie, J., Gan, M. & Brooks, C. (2017). Instruction Based on Feedback. In R. E. Mayer & P. A. Alexander (Hrsg.), *Handbook of Research on Learning and Instruction* (2nd ed.) (S. 290-324). New York: Routledge.

Häcker, H. Leutner, D. & Amelang, M. (2013). Beurteilen und Bewerten schulischer Leistungen. In M. Hasselhorn, H. Hauer & S. Schneider (Hrsg.), *Pädagogische Psychologie: Erfolgreiches Lernen und Lehren*, S. 389-409. Stuttgart: Kohlhammer

Häcker, H. Leutner, D. & Amelang, M. (1998). Standards für pädagogisches und *psychologisches Testen*. Göttingen: Hogrefe.

Harks, B., Rakoczy, K., Hattie, J., Besser, M. & Klieme, E. (2014). The effects of feedback on achievement, interest and self-evaluation: The role of feedback's perceived usefulness. *Educational Psychology*, 34 (3), S. 269–290.

Hartinger, A., Fölling-Albers, M. & Mörtl-Hafizović, D. (2005). Die Bedeutung der Ambiguitätstoleranz für das Lernen in situiereten Lernbedingungen. *Psychologie in Erziehung und Unterricht*.

Harwell, M. R. & Gatti, G. G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research*, 71(1), S. 105-131.

Havighurst, R. J. (1972). *Developmental Tasks and Education*. New York: McKay.

Hascher, T. & Edlinger, H. (2009). Positive Emotionen und Wohlbefinden in der Schule – Ein Überblick über Forschungszugänge und Erkenntnisse. *Psychologie in Erziehung und Unterricht*, 56 (2), S. 105-122.

Heid, H. (1999). Über die Vereinbarkeit individueller Bildungsbedürfnisse und betrieblicher Qualifikationsanforderungen. *Zeitschrift für Pädagogik*, 2, S. 231-244.

Heller, K. A. (1997). Individuelle Bedingungsfaktoren der Schulleistung: Literaturüberblick. In F.E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter*. Weinheim: Beltz. S. 183–201.

Helmke, A. (1993). Die Entwicklung der Lernfreude vom Kindergarten bis zur 5. Klassenstufe. *Zeitschrift für Pädagogische Psychologie*, 7 (2/3), 1993, S. 77–86.

Helmke, A. & Weinert, F. E. (1997): Bedingungsfaktoren schulischer Leistungen. In: Weinert, F. E. (Hrsg.): *Psychologie des Unterrichts und der Schule – Enzyklopädie der Psychologie, Pädagogische Psychologie*, Bd. 3 – Göttingen, S. 71–176.

Helmke, A. & Weinert, F. E. (1999). Schooling and the development of achievement differences. In: Weinert, F. E. & Schneider, W. (Hrsg.): *Individual Development from 3 to 12: Findings from the Munich Longitudinal Study*. New York, S. 176–192.

Hoyt, W. T. & Kerns, M. D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods*, 4, S. 403–424. Ingenkamp, K. (Hrsg.) (1995): *Die Fragwürdigkeit der Zensurengebung. Texte und Untersuchungsberichte*. Weinheim: Basel.

Hess, M. (2019). Professionell Feedback geben – Lernen mit Videos in der ersten Phase der Grundschullehrerbildung. In *Grundschulpädagogik zwischen Wissenschaft und Transfer* (S. 102-107). Springer VS, Wiesbaden.

Jürgens, E., & Sacher, W. (2008). *Leistungserziehung und pädagogische Diagnostik in der Schule: Grundlagen und Anregungen für die Praxis*. W. Kohlhammer Verlag.

Vergleichsstudie (2003). *Vertiefender Vergleich der Schulsysteme ausgewählter PISA-Teilnehmerstaaten: Kanada, England, Finnland, Frankreich, Niederlande, Schweden*. BMBF, Referat Öffentlichkeitsarbeit.

Lintorf, K. (2011). *Wie vorhersagbar sind Grundschulnoten?: Prädiktionskraft individueller und kontextspezifischer Merkmale*. Springer-Verlag.

Lohbeck, A., Hagenauer, G., & Moschner, B. (2016). Zum Zusammenspiel zwischen schulischem Selbstkonzept, Lernfreude, Konzentration und Schulleistungen im Grundschulalter. *Zeitschrift Für Bildungsforschung*, 6(1), S. 53–69. <https://doi.org/10.1007/s35834-016-0147-2>

Lüdtke, O., Köller, O., Marsh, H. W., & Trautwein, U. (2005). Teacher frame of reference and the big-fish–little-pond effect. *Contemporary educational psychology*, 30(3), 263–285.

Lütgert, W., Tillmann, K. J., Beutel, S. I., Jachmann, M., & Vollstädt, W. (2001). Leistungsbeurteilung und Leistungsrückmeldung an Hamburger Schulen. *Bericht über ein Forschungsprojekt*.

Mayr, J. (2003). Persönlichkeitsfragebögen in der Lehrerforschung und Lehrerberatung. In K. Samac (Hrsg.), *Empirisches Arbeiten in der Arbeitsgemeinschaft der Bewegungserzieherinnen und Bewegungserzieher an Pädagogischen Akademien*. Wien: BMBWK.

Mayr, J. (2009). LehrerIn werden in Österreich: empirische Befunde zum Lehramtsstudium. *Erziehung und Unterricht* 159(1/2), S. 14 – 33.

Malouff, J. M., & Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education*, 60(3), 245–256. <https://doi.org/10.1177/0004944116664618>

Marsh, H.W. (1987): The big-fish-little-pond effect on academic self-concept. In: *Journal of Educational Psychology*. 79. S. 280-295.

Marsh, H.W., Kuyper, H., Morin, A.J.S., Parker, P.D. & Seaton, M. (2014). Big-fish-little-pond social comparison and local dominance effects: Integrating new statistical models, methodology, design, theory and substantive implications. *Learning and Instruction*, 33, S. 50–66.

McCrae, R. R., & Costa, P. T. (1997). Conceptions and Correlates of Openness to Experience. *Handbook of Personality Psychology*, (November), S. 825–847. <https://doi.org/10.1016/b978-012134645-4/50032-9>

Meece, J. L., Anderman, E. M., & Anderman, L. H. (2006). Classroom goal structure, student motivation, and academic achievement. *Annual Review of Psychology*, 57, S. 487-503.

McKown, C., & Weinstein, R. S. (2003). The development and consequences of stereotype consciousness in middle childhood. *Child development*, 74(2), 498-515.

Mischo, C., & Rheinberg, F. (1995). Erziehungsziele von Lehrern und individuelle Bezugsnormen der Leistungsbewertung. *Zeitschrift für pädagogische Psychologie*, 9(3), 139-151.

Mohan, G., & Mulla, Z. R. (2013). Openness to experience and work outcomes: exploring the moderating effects of conscientiousness and job complexity. *Great Lakes Herald*, 7(2), S. 18-36

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, 79(3), 1129-1167.

Murphy, K. R., & Anhalt, R. L. (1992). Is halo error a property of the rater, ratees, or the specific behaviors observed? *Journal of Applied Psychology*, 77, S. 494–500.

Neuweg, G. (2014). *Schulische Leistungsbeurteilung. Rechtliche Grundlagen und pädagogische Hilfestellungen für die Schulpraxis*. Linz: Trauner Verlag.

Nisbett R. E. & DeCamp Wilson T. (1977). The Halo Effect: Evidence for unconscious alteration of judgements. *Journal of Personality and Social Psychology*, Vol. 35(4), S. 250-256. <https://doi.org/10.1037/0022-3514.35.4.250>

Noor, N. M., Aman, I., Mustafa, R., & Seong, T. K. (2010). Teacher's verbal feedback on students' response: A Malaysian ESL classroom discourse analysis. *Procedia - Social and Behavioral Sciences*, 7(2), S. 398–405. <https://doi.org/10.1016/j.sbspro.2010.10.054>

Nunn, R. (2001). 'Language learning across boundaries-Negotiating Classroom Rituals.' Zugriff am 20. Dezember 2019. Verfügbar unter <http://www-writing.berkeley.edu/TESL-EJ/ej18/al.html>

OECD. (2019). *PISA 2018 Results: What Student Know and Can Do*. <https://doi.org/10.1787/5f07c754-en>

Oerter, R. (1998). Kindheit. In R. Oerter & L. Montada, *Entwicklungspsychologie. Ein Lehrbuch* (4. Aufl.). Weinheim: Beltz Psychologie Verlags Union.

Paasch, D., Schmid, C., Kallinger-Aufner, A., & Knollmüller, R. (2019). Noten und Kompetenzen in verschiedenen Fächern, Schulstufen und Schulformen. In *Kompetenzmessungen im österreichischen Schulsystem: Analysen, Methoden & Perspektiven*, S. 161-177.

Pekrun, R. (1997). Selbstkonzept und Leistung – Dynamik ihres Zusammenspiels: Kommentar zu Möglichkeiten und Grenzen differentialpsychologischer Analysen. In: F.E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter*, S. 351-358. Weinheim: Beltz.

Rakoczy, K., Pinger, P., Hochweber, J., Klieme, E., Schütze, B., & Besser, M. (2019). Formative assessment in mathematics: Mediated by feedback's perceived usefulness and students' self-efficacy. *Learning and Instruction*, 60, S. 154-165.

Rechnungshof (2015). *Bericht des Rechnungshofes zu den Schulversuchen im Schuljahr 2012/13*

- Rheinberg, F. (1980). *Leistungsbewertung und Lernmotivation*. Göttingen: Hogrefe.
- Rheinberg, F. (2006). Bezugsnormorientierung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie*, 3. Aufl., S. 55–62. Weinheim: Beltz, PVU.
- Rheinberg, F., Bromme, R., Minsel, B., Winteler, A., & Weidenmann, B. (2001). Die Erziehenden und Lehrenden. *Pädagogische Psychologie*, 4, S. 271-355.
- Ruiz-Primo, M. A. & Brookhart, S. M. (2017). *Using Feedback to Improve Learning*. New York: Routledge.
- Salchegger, S. (2016). Selective school systems and academic self-concept: How explicit and implicit school-level tracking relate to the big-fish–little-pond effect across cultures. *Journal of Educational Psychology*, 108, 405–423. doi:10.1037/edu0000063
- Satow, L. (2012). *Big-Five-Persönlichkeitstest (B5T): Testmanual und Normen*.
- Satow, L. (2011). Big-Five-Persönlichkeitstest (B5T) [Testbeschreibung und Items]. In Zentrum für Psychologische Information und Dokumentation (ZPID) (Hrsg.), PSYINDEX (Online-Datenbanksegment PSYINDEX Tests Dok.-Nr. 9006357). Trier: ZPID.
- Schmid, C., Paasch, D., & Katstaller, M. (2016). Kompositionseffekte bei der Notenvergabe in Mathematik auf der 4.Schulstufe der österreichischen Volksschule. *Zeitschrift Für Bildungsforschung*, 6(3), S. 265–283. <https://doi.org/10.1007/s35834-016-0170-3>
- Schmidt, F.L. & Hunter, J. E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin*, 124 (2), S. 262-274.
- Schmude, C. (2005). *Differenzielle Entwicklungsverläufe der Lernfreude im Grundschulalter*.
- Schneider, W. (Hrsg.) (2008). *Entwicklung von der Kindheit bis zum Erwachsenenalter: Befunde der Münchner Längsschnittstudie*. Weinheim: Beltz Psychologie Verlags Union.

Schneider, W., & Bullock, M. (2008). *Die Längsschnittstudie LOGIK: Versuch einer zusammenfassenden Würdigung*. Entwicklung von der Kindheit bis zum Erwachsenenalter. Befunde der Münchner Längsschnittstudie LOGIK, S. 203-218. Weinheim, Basel: Beltz.

Schreiner, C., Breit, S., Pointinger, M., Pacher, K., Neubacher, M. & Wiesner, C. (Hrsg.). (2018). *Standardüberprüfung 2017. Mathematik, 8. Schulstufe. Bundesergebnisbericht*. S. 33 f. Salzburg: BIFIE. Verfügbar unter <https://www.bifie.at/material/ueberpruefung-der-bildungsstandards/ergebnisberichte/>

Schuler, H. (2005). *Lehrbuch der Personalpsychologie*. Göttingen: Hogrefe.

Silvia, P.J., Beaty, R.E., Nusbaum, E.C., Eddington, K.M., Levin-Aspenson, H., & Kwapil, T.R. (2014). Everyday creativity in daily life: An experience-sampling study of “little c” creativity. *Psychology of Aesthetics, Creativity, and the Arts*, 8, S. 183–188; doi:10.1037/a0035722.

Smith, E., & Gorard, S. (2005). “They don’t give us our marks”: The role of formative feedback in student progress. *Assessment in Education: Principles, Policy & Practice*, 12(1), 21–38. <https://doi.org/10.1080/0969594042000333896>

Sorrentino, R. M., & Roney, C. J. R. (2000). *Essays in social psychology*.

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. *Annual review of psychology*, 67, S. 415-437.

Spiel, C. (1997). Das Stereotyp des schlechten Schülers: Diskussion. *Entwicklung im Grundschulalter, Weinheim: Psychologie Verlags Union*, S. 281-286.

Spiel, C. & Wagner, P. (2003). Funktion von Schule und Bewertungen durch Schule. *Erziehung und Unterricht*, 153(5-6), S. 643-654.

Spiel, C., Schober, B., Wagner, P. & Reimann, R. (Hrsg.) (2010). *Bildungspsychologie*. S. 83. Göttingen: Hogrefe.

Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19, S. 85-95.

Spinath, B. (2010). Bildungspsychologie des Primärbereichs. In C. Spiel, B. Schober, P. Wagner & R. Reimann (Hrsg.), *Bildungspsychologie*, S. 80-97. Göttingen: Hogrefe.

Spinath, B., & Spinath, F. M. (2005). Development of self-perceived ability in elementary school: the role of parents' perceptions, teacher evaluations, and intelligence. *Cognitive Development*, 20(2), S. 190–204. <https://doi.org/10.1016/J.COGDEV.2005.01.001>

Sprietsma, M. (2013). Discrimination in grading: experimental evidence from primary school teachers. *Empirical Economics* 45(1), S. 523-538.

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological review*, 115(2), S. 336

Stern, T. (2010). *Förderliche Leistungsbewertung. Österreichisches Zentrum für Persönlichkeitsbildung und soziales Lernen (ÖZEPS)*. Wien: Bundesministerium für Unterricht, Kunst und Kultur (BMUKK).

Stobart, G. & Gipps, C. (1997) *Assessment: a teacher's guide to the issues*. London, Hodder and Stoughton.

Stigler, J., & Hiebert, J. (1997). Understanding and improving classroom mathematics instruction: An overview of the TIMSS video study. In *ACER National Conference 1997*, S. 52-65.

Tan, C. S., Lau, X. S., Kung, Y. T., & Kailsan, R. A. (2019). Openness to Experience Enhances Creativity: The Mediating Role of Intrinsic Motivation and the Creative Process Engagement. *Journal of Creative Behavior*, 53(1), 109–119. <https://doi.org/10.1002/jocb.170>

Tatalović Vorkapić, S. (2012). The Significance of preschool teacher's personality in early childhood education: Analysis of Eysenck's and Big Five dimensions of personality.

*International Journal of Psychology and Behavioral Sciences*, 2(2), 28–37.  
<https://doi.org/10.5923/j.ijpbs.20120202.05>

Tent, L. & Birkel P. (2010). Zensuren. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 949–958). Weinheim, Basel: Beltz.

Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of applied psychology*, 4(1), S. 25-29.

Tillmann, K.-J. & Vollstädt, W. (2000). Ziffernote versus Lernbericht. Funktionen der Leistungsbeurteilung in unterschiedlichen Schulstufen und Bildungsgängen – Eine schultheoretische Einordnung. In M. Sertl, K.-J. Tillmann & W. Vollstädt (Hrsg.), *Noten – nicht zu umgehen? Alternative Formen der Leistungsbeurteilung auf dem Prüfstand* (S. 30–60). Wien: Verein der Förderer der Schulhefte.

TIMSS, I. (1995). IEA's Third International Mathematics and Science Study. TIMSS Science Items: Released Set for Population 2 (Seventh and Eighth Grades). *Chestnut Hill, MA: Boston College. This paper uses data from the National Educational Panel Study (NEPS): Additional Study Thuringia*

Valtin, R., Schmude, C., Rosenfeld, H., Darge, K., Ostrup, G., Thiel, O., Wagener, M. & Wagner, C. (2002). *Was ist ein gutes Zeugnis? Noten und verbale Beurteilungen auf dem Prüfstand*. Weinheim: München: Juventa Verlag.

Valtin, R. & Rosenfeld, H. (1997). *Zur Präferenz von Noten- oder Verbalbeurteilungen – Ein Vergleich Ost- und Westberliner Eltern*. In: H.E. Elmar (Hrsg.), *Zeitschrift für Pädagogik, Beiheft*, 37, S. 293-304. Weinheim: Beltz.

Voerman, L., Meijer, P. C., Korthagen, F. A.J. & Simons, R. J. (2012). Types and Frequencies of Feedback Interventions in Classroom Interaction in Secondary Education. *Teaching and Teacher Education*, 28(8), S. 1107-1115.

Wagener, M. (2003). Ziffernzensuren oder verbale Beurteilung. *Eine empirische Untersuchung zum Zusammenhang von Beurteilungsform und Unterrichtsorganisation, Leistungsrückmeldung und Klassenraumgestaltung von Grundschullehrerinnen aus östlichen und westlichen Bezirken Berlins*. Weinheim: Beltz Verlag.

Weber, H. M., & Petermann, F. (2016). Der Zusammenhang zwischen Schulangst, Schulunlust, Anstrengungsvermeidung und den Schulnoten in den Fächern Mathematik und Deutsch. *Zeitschrift für Pädagogik*, 62(4), S. 551-570.

Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessung in Schulen*, S. 17-31. Weinheim: Beltz

Weinert, F. E. & Helmke, A. (Hrsg.) (1997) *Entwicklung im Grundschulalter*. Weinheim: Beltz.

William, D. (2013). Assessment: The bridge between teaching and learning. *Voices from the Middle*, 21(2), S. 15.

William, D. (2014). Formative assessment and contingency in the regulation of learning processes. In *Annual Meeting of American Educational Research Association*, Philadelphia, PA.

Winter, F. (2015). *Lerndialog statt Noten: neue Formen der Leistungsbeurteilung; [Verbalzeugnis, Beurteilungsraster, Portfolio, Lernentwicklungsgespräch]*. Beltz.

Winter, F. (2004). *Leistungsbewertung. Eine neue Lernkultur braucht einen anderen Umgang mit den Schülerleistungen*. Baltmannsweiler: Schneider.

Yerkes, R. M. & Dodson, J.D. (1908). The Relation of Strength of Stimulus to Rapidity of Habitformation. *Journal of Comparative Neurology of Psychology*, 18, S. 482-495.

## 12. Anhang

### Anhang A: Zusammenfassung

In Österreich wurde im Schuljahr 2019/20 die Beurteilung durch Ziffernnoten in Volksschulen ab dem Jahreszeugnis der zweiten Klasse wieder eingeführt, wobei die jeweiligen Eltern, über die Art der Leistungsbeurteilung, bis zum Ende des ersten Semesters der zweiten Schulklasse, mitbestimmen können. Deshalb werden sowohl Lehrpersonen als auch Laien in die Untersuchung miteinbezogen. Das Ziel dieser Arbeit war es, die Leistungsbeurteilungspräferenz Verbalbeurteilungen versus Ziffernnoten zu untersuchen.

*Offenheit (Big-Five)* wird in der Literatur mit der Vorliebe für komplexe Herausforderungen und individualisierte Lehrmethoden in Verbindung gebracht. Weiters gibt es Evidenz für die tendenzielle Beurteilungsverzerrung (*Halo-Effekt*) bei der Leistungsbeurteilung. In dieser Studie wurden Laien und Lehrpersonen hinsichtlich ihrer Leistungsbeurteilungspräferenz je Volksschulklasse, der Zusammenhang von *Offenheit* und der Präferenz für Verbalbeurteilungen je Klasse sowie die Beurteilungsverzerrung der Benotung eines fiktiven Aufsatzes durch irrelevante Nebeninformationen untersucht. Die Erhebung fand mittels Online-Fragebogen statt.

Die Ergebnisse zeigten keinen Unterschied in der Beurteilungspräferenz zwischen Lehrer\*innen und Laien. Es wurden Zusammenhänge zwischen *Offenheit* und der Leistungsbeurteilungspräferenz für Verbalbeurteilungen in jeder der vier Schulklassen gefunden. Nachweise für einen *Halo-Effekt* betreffend die Benotung der fiktiven Geschichte konnten nicht gefunden werden. Es wurden jedoch große Unterschiede bei den Benotungen durch die Lehrer\*innen entdeckt. Diese und weitere explorative Ergebnisse dieser Untersuchung sprechen für einen Konsens der bereits in der Literatur vorwiegenden Bevorzugung von verbalen Leistungsbeurteilungen gegenüber Ziffernnoten und der schlechten Gütekriterien von Schulnoten. Die möglichen Hintergründe der Ergebnisse und Einschränkungen dieser Untersuchung werden diskutiert.

## Anhang B: Abstract

In Austria, the assessment by numerical grades in primary schools was reintroduced in the school year 2019/20 from the second grade annual certificate, whereby the respective parents can have a say in the type of performance assessment until the end of the first semester of the second grade. Therefore, both teachers and laypersons are included in the study. The aim of this work was to examine the performance assessment preference of verbal assessments versus numerical grades.

In the literature, *openness (big-five)* is associated with a preference for complex challenges and individualized teaching methods. Furthermore, there is evidence for the tendency to bias (*halo effect*) in performance assessment. In this study laypersons and teachers were examined with regard to their performance assessment preference per primary school class, the connection between *openness* and the preference for verbal assessment per class as well as the bias in the assessment of the rating of a fictional essay due to irrelevant secondary information. The survey was carried out using an online questionnaire.

The results showed no difference in the assessment preference between teachers and laypersons. Relationships between *openness* and the preference for verbal appraisals were found in each of the four school classes. Evidence for a *halo-effect* regarding the grading of the fictional story could not be found. However, great differences in the grades by the teachers were discovered. These and other exploratory results of this study speak for a consensus of the preference in the literature for verbal performance assessments over numerical grades and the poor-quality criteria for school grades. The possible background of the results and limitations of this investigation are discussed.

## Anhang C: Rekrutierungsmail Lehrpersonen (Volksschulen)

Sehr geehrte Damen und Herren!

im Rahmen meiner Masterarbeit an der Fakultät für Psychologie der Universität Wien, führe ich in Zusammenarbeit mit meiner Betreuerin, Frau Univ.-Prof. DDr. Christiane Spiel, eine Studie zum Thema „Leistungsbeurteilung in Volksschulen“ durch.

Deshalb melde ich mich bei Ihnen in der Hoffnung, Sie könnten mich bei dieser Aufgabe unterstützen und den von mir erstellten Fragebogen beantworten sowie an Ihre Mitarbeiter\*innen weiterleiten. Die Studie richtet sich neben Direktor\*innen und Lehrer\*innen auch an sonstige volljährige Personen.

Der Link zum Fragebogen lautet:

<https://www.soscisurvey.de/volksschule-leistungsbeurteilung/>

Eine Durchführung ist gleichermaßen auf allen Eingabegeräten möglich.

Die Bearbeitung benötigt ca. 10 Minuten.

Die Antworten werden anonym und vertraulich behandelt, sodass diese nicht auf die jeweilige Person zurückführbar sind.

Mit Ihrer Teilnahme tragen Sie dazu bei, bildungspsychologische Prozesse besser zu verstehen.

Bei anfallenden Fragen stehe ich Ihnen gerne unter der E-Mail-Adresse [a00761531@unet.univie.ac.at](mailto:a00761531@unet.univie.ac.at) zur Verfügung.

Vielen herzlichen Dank,

Philipp Wurmitzer, BSc.

## Anhang D: Rekrutierungsschreiben Laien (Online-Foren)

Sehr geehrte Damen und Herren!

im Rahmen meiner Masterarbeit an der Fakultät für Psychologie der Universität Wien, führe ich in Zusammenarbeit mit meiner Betreuerin, Frau Univ.-Prof. DDr. Christiane Spiel, eine Studie zum Thema „Leistungsbeurteilung in Volksschulen“ durch.

Deshalb melde ich mich bei Ihnen in der Hoffnung, Sie könnten mich bei dieser Aufgabe unterstützen und den von mir erstellten Fragebogen beantworten. Die Studie richtet sich an alle volljährigen Personen.

Der Link zum Fragebogen lautet:

<https://www.soscsurvey.de/volksschule-leistungsbeurteilung/>

Eine Durchführung ist gleichermaßen auf allen Eingabegeräten möglich.

Die Bearbeitung benötigt ca. 5 bis 10 Minuten.

Die Antworten werden anonym und vertraulich behandelt, sodass diese nicht auf die jeweilige Person zurückführbar sind.

Mit Ihrer Teilnahme tragen Sie dazu bei, bildungspsychologische Prozesse besser zu verstehen.

Bei anfallenden Fragen stehe ich Ihnen gerne unter der E-Mail-Adresse [a00761531@unet.univie.ac.at](mailto:a00761531@unet.univie.ac.at) zur Verfügung.

Vielen herzlichen Dank,

Philipp Wurmitzer, BSc.

# Anhang E: Online-Fragebogen



## **Liebe Teilnehmerinnen und Teilnehmer,**

**Startseite**

im Rahmen meiner Masterarbeit an der Fakultät für Psychologie der Universität Wien führe ich eine Studie zum Thema "Leistungsbeurteilung in der Volksschule" durch.

Mit Ihrer Teilnahme ermöglichen Sie, bildungspsychologische Prozesse besser zu verstehen.

Die Bearbeitung benötigt ungefähr 5 bis 10 Minuten.

Bitte beantworten Sie die nachfolgenden Fragen wahrheitsgemäß und nach Möglichkeit vollständig. Es gibt keine falschen oder richtigen Antworten.

Ihre Antworten werden anonym und vertraulich behandelt, sodass diese nicht auf Ihre Person zurückführbar sind.

**Danke für Ihre Mitarbeit!**

Bitte geben Sie Ihr Geschlecht an

SD01 

- weiblich
- männlich
- anderes

Bitte geben Sie Ihr Alter an

SD02 

Jahre

Was ist Ihr höchster absolvierter Bildungsabschluss?

SD03 

- Pflichtschulabschluss
- Lehre (mit Lehrabschlussprüfung)
- berufsbildende mittlere Schule
- Hochschulreife (Matura, Studienberechtigung, etc.)
- Bachelor
- Master/Diplom
- Doktor/PhD
- Pädagogische Hochschule/Pädagogische Akademie
- anderer Bildungsabschluss:

Haben Sie Kinder?

SD04 

- Ja, mindestens eines besucht die Volksschule
- Ja, keines davon besucht die Volksschule
- Nein

Befinden Sie sich derzeit in einem aufrechten Arbeitsverhältnis als Volksschullehrer\*in?

SD05 

- Ja
- Nein, aber ich war als Volksschullehrer\*in beschäftigt
- Nein, ich befinde mich aktuell in der Ausbildung zur/ zum Volksschullehrer\*in
- Nein

Befinden Sie sich derzeit in einem aufrechten Arbeitsverhältnis als Volksschuldirektor\*in?

SD06 

- Ja
- Nein, aber ich war als Volksschuldirektor\*in beschäftigt
- Nein

**2 aktive(r) Filter**

**Filter SD06/F1**

Wenn eine der folgenden Antwortoption(en) ausgewählt wurde: 1, 2  
Dann Frage/Text LE02 später im Fragebogen anzeigen (sonst ausblenden)

**Filter SD06/F2**

Wenn eine der folgenden Antwortoption(en) ausgewählt wurde: 1, 2  
Dann Frage/Text LE04 später im Fragebogen anzeigen (sonst ausblenden)

---

**Seite 03**

LEjaFilter

**PHP-Code**

```
if ((value('SD05') == 1) or (value('SD05') == 2))
{
    goToPage ('LEja');
}
elseif ((value('SD06') == 1) or (value('SD06') == 2))
{
    goToPage ('LEja');
}
elseif ((value('SD05') == 3))
{
    goToPage ('GE');
}
else
{
    goToPage ('B5T');
}
```

Bitte geben Sie in Summe an, wie lange Sie bereits als Volksschuldirektor\*in tätig waren:

LE02 

Bei einer Beschäftigungszeit von unter einem Jahr, bitte auf 1 aufrunden.

 Jahr(e)

Bitte geben Sie in Summe an, wie lange Sie bereits als Volksschullehrer\*in tätig waren:

LE01 

Bei einer Beschäftigungszeit von unter einem Jahr, bitte auf 1 aufrunden.

 Jahr(e)

In wie vielen Volksschulen waren Sie bisher als Direktor\*in beschäftigt?

LE04  Schule(n)

In wie vielen Volksschulen waren Sie bisher als Lehrer\*in beschäftigt?

LE03  Schule(n)

Bitte geben Sie an, in welchem Wiener Bezirk oder Bundesland sich die Schule befindet, in der Sie aktuell angestellt  
zuletzt angestellt waren:

LE05  ▼

Absolvierten Sie nach Ihrer Ausbildung zur Volksschullehrer\*in (eine) Weiterbildung/en, im Bereich der Leistungsbeurteilung  
von Volksschüler\*innen?

LE06 

- Ja
- Nein, ich absolviere aktuell eine Weiterbildung im Bereich der Leistungsbeurteilung
- Nein, ich habe vor, demnächst eine Weiterbildung im Bereich der Leistungsbeurteilung zu besuchen
- Nein

Haben Sie in der Beurteilung von Schülerleistungen mehr praktische Erfahrung mit Ziffernnoten- oder verbaler  
Leistungsbeurteilungen?

LE07 

- Verbalbeurteilungen
- Ziffernnotenbeurteilung
- ungefähr gleich viel

---

kann ich nicht beurteilen

Unterrichten/Unterrichteten Sie das Schulfach „Deutsch“?

LE08 

- ja
- nein

### **Alternative 1 von 3 zu Peters Geschichte (Seite 5a):**

---

Seite 05  
GE

#### **Peters Geschichte**

GE01

Stellen Sie sich vor, Sie sind Klassenlehrer\*in einer 4. Klasse einer öffentlichen Volksschule. Sie sind gerade dabei, eine Aufgabe zu verbessern und zu benoten. Das Thema zur Geschichte, die die Kinder selbständig verfassen sollten, war "Erzähle ein Erlebnis, wo du Angst hattest". Sie verbessern gerade die Geschichte von Peter. Er ist 10 Jahre alt und kommt immer ordentlich gekleidet und ausgeschlafen in die Schule. Peter war seit Beginn der ersten Klasse einer Ihrer fleißigsten Schüler. Nie hat er eine Hausaufgabe vergessen zu erledigen und im Unterricht passte er auch immer auf.

Bitte klicken Sie auf "weiter", um mit der Benotung von Peters Geschichte fortzufahren.

---

### **Alternative 2 von 3 zu Peters Geschichte (Seite 5b):**

---

Seite 05  
GE

#### **Peters Geschichte**

GE02

Stellen Sie sich vor, Sie sind Klassenlehrer\*in einer 4. Klasse einer öffentlichen Volksschule. Sie sind gerade dabei, eine Aufgabe zu verbessern und zu benoten. Das Thema zur Geschichte, die die Kinder selbständig verfassen sollten, war "Erzähle ein Erlebnis, wo du Angst hattest". Sie verbessern gerade die Geschichte von Peter. Er ist 10 Jahre alt und kommt meist unordentlich gekleidet und verschlafen in die Schule. Er war seit Beginn der ersten Klasse einer Ihrer faulsten Schüler. Er hat bisher viele Hausaufgaben vergessen zu erledigen und im Unterricht passte er auch sehr selten auf.

Bitte klicken Sie auf "weiter", um mit der Benotung von Peters Geschichte fortzufahren.

---

### **Alternative 3 von 3 zu Peters Geschichte (Seite 5c):**

## Peters Geschichte

GE03

Stellen Sie sich vor, Sie sind Klassenlehrer\*in einer 4. Klasse einer öffentlichen Volksschule. Sie sind gerade dabei, eine Aufgabe zu verbessern und zu benoten. Das Thema zur Geschichte, die die Kinder selbständig verfassen sollten, war "Erzähle ein Erlebnis, wo du Angst hattest". Sie verbessern gerade die Geschichte von Peter. Er ist 10 Jahre alt.

Bitte klicken Sie auf "weiter", um mit der Benotung von Peters Geschichte fortzufahren.

### „Der Tiger im Bett“

GE04

Eines Mittwochs des Dezembers schimpfte meine Mutter wieder einmal. Mein Bruder hatte noch Hausübung und ich wollte nicht ins Bett. Schließlich gab ich doch nach und ging ganz langsam hinauf. Als ich in mein Zimmer ging, sah ich etwas auf meinem Bett liegen. Wenn ich genauer hinschaute, sag es so aus als wäre es ein liegender Tiger. Ich wurde blass, totenblass. Im Zimmer war es nämlich stockfinster, und ich konnte nur wenig erkennen. Ich wollte mich ducken, doch da stis ich an. Ich sah aus den Augenwinckeln, dass sich der Tiger bewegte, Und ich schrie „Hilfe“. Das Herzklopfte mir bis zum Hals. Ich hatte angst, Todesagst und ich dachte dass er mich fressen würde. Ich rannte zum Lichtschalter und schaltete das Licht an. Als ich sah, dass der Tiger nicht anderes war als Decken und Polster, war Ich sehr erleichtert. Die Ohren waren die Sofapolster und der Körper die Bettdecke. Als meine Mutter endlich kam, Erzelte ich ihr alles und leider hatte ich seit dem Agst vor Tigern und anderenen Wildtieren.

Falls Sie, für sich selbst, Notizen zur Benotung machen wollen, können Sie dafür dieses Eingabefeld nutzen:

GE05

Bitte beurteilen Sie Peters Geschichte mit einer Schulnote

1	2	3	4	5	kann ich nicht beurteilen
---	---	---	---	---	---------------------------

### PHP-Code

```
if (value('GE05_01') == -1) {  
    goToPage('GE4');  
} else {  
    goToPage('B5T');  
}
```

Bitte begründen Sie mit ein paar Worten, warum Sie Peters Geschichte nicht benoteten.

GE06

keine Angabe

Bitte beantworten Sie die folgenden Fragen so zutreffend wie möglich. Es gibt keine falschen Antworten.

OF01

	trifft gar nicht zu	trifft eher nicht zu	trifft eher zu	trifft genau zu
Ich will immer neue Dinge ausprobieren.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich bin ein neugieriger Mensch.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich reise viel, um andere Kulturen kennenzulernen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Am liebsten ist es mir, wenn alles so bleibt, wie es ist.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich diskutiere gerne.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich lerne immer wieder gerne neue Dinge.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich beschäftige mich viel mit Kunst, Musik und Literatur.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich interessiere mich sehr für philosophische Fragen.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich lese viel über wissenschaftliche Themen, neue Entdeckungen oder historische Begebenheiten.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Ich habe viele Ideen und viel Fantasie.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Geben Sie an, ob Ihrer Meinung nach Schulleistungen von Volksschüler\*innen mittels Verbalbeurteilungen oder Ziffernoten von 1 bis 5 angemessener beurteilt werden können.

BE01

	Verbal beurt.	Ziffern noten	kann ich nicht beurteilen
1. Klasse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2. Klasse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
3. Klasse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
4. Klasse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Wenn Sie Ihre Präferenzen zur Leistungsbeurteilung begründen wollen, können Sie dieses Texteingabefeld dafür nutzen.

BE02

maximal 240 Zeichen

## Vielen Dank für Ihre Teilnahme!

Sollten Sie Interesse an den Forschungsergebnissen dieser Studie haben, oder etwaige Fragen auftreten, melden Sie sich bitte unter folgender E-Mail-Adresse: a00761531@unet.univie.ac.at

## Anhang F: Items-Offenheit (B5T, Satow, 2012)

Die Items wurden durch eine Faktorenanalyse und anschließender Varimax-Rotation ausgewählt. Nur die geeignetsten Items, die hoch auf den Faktor *Offenheit* luden, wurden ausgewählt und danach mit Scree-Test bestätigt (Satow, 2012). Es folgen die 10 ausgewählten und für die Untersuchung verwendeten Items mit den wichtigsten Kennwerten:

Tabelle 11: Veranschaulichung der Items zur Kategorie „Offenheit“:

Item	Mittelwert	SD	Trennschärfe
1) Ich will immer neue Dinge ausprobieren.	3.0	0.75	0.59
2) Ich bin ein neugieriger Mensch.	3.4	0.64	0.58
3) Ich reise viel, um andere Kulturen kennenzulernen.	2.5	0.82	0.39
4) Am liebsten ist es mir, wenn alles so bleibt, wie es ist.* (-)	2.8	0.82	0.31
5) Ich diskutiere gerne.	3.1	0.84	0.46
6) Ich lerne immer wieder gerne neue Dinge.	3.4	0.63	0.60
7) Ich beschäftige mich viel mit Kunst, Musik und Literatur.	2.7	0.97	0.45
8) Ich interessiere mich sehr für philosophische Fragen.	2.9	0.96	0.49
9) Ich lese viel über wissenschaftliche Themen, neue Entdeckungen oder historische Begebenheiten.	2.8	1.00	0.46
10) Ich habe viele Ideen und viel Fantasie.	3.2	0.75	0.55

Tabelle 11: Fragebogen zur Erfassung der *Offenheit* (Satow, 2012). \*Bei dem negativ gepolten Item 4 verlief die Punktevergabe von 4 bis 1.

## Anhang G: Begründungen für die Beurteilungspräferenzen (Lehrer\*innen)

Tabelle 12: Begründungen für die Angaben der Präferenzen der Beurteilungsmethoden in den vier Volksschulklassen der Proband\*innen die als Lehrer\*innen beschäftigt waren (n = 46):

1	2	3	4	Begründung
V	V	V	V	Mit alternativen Beurteilungsformen kann man Eltern und Kinder viel differenzierter über den Leistungsstand, über Stärken und Schwächen informieren und genau aufzeigen, wo noch Übungsbedarf besteht. 
V	V	Z	Z	1./2.Kl. verb. Beurteilung gibt Raum, um Schwächen/Talente ausführlicher und detaillierter anzusprechen; 3./4. Ziffernbeurt. als Hinführung auf die Ziffernbenotung in den nächsthöheren Schultypen
V	V	Z	Z	Ab der 3. Klasse Ziffernnoten UND verbale Beurteilung.
V	V	Z	Z	Am besten finde ich das gemeinsame Gespräch mit Eltern und Kind, die sogenannte Lernfortschrittsdokumentation (LFD), mit einem kindgemäßen schriftlichen Feedback (Was kann ich, was muss ich noch üben).
-	-	-	-	Angesichts der unterschiedlichen Start-, Lebens- und Entwicklungsbedingungen gerade in den ersten Schuljahren der Kinder würde ich mir sehr viel mehr Differenzierung und Förderung wünschen. Dem gegenüber sind Ziffernnoten brachial
Z	Z	Z	Z	Blickrichtung weiterführende Schulen
V	V	V	Z	Das "Bemühen" kann in einer Ziffernnote nicht dargestellt werden!!!
V	V	V	V	Der Lernfortschritt und erreichte Lernziele können beschrieben werden.
V	Z	Z	Z	Die Kinder wollen..verbale kann Ergänzung sein..meist sind es ohnedies Standardsätze, wie bei einer Dienstbeurteilung. Schlimm sind in Wien die Migranten und Asylanten. Die machen unser System kaputt. So ist der Druck der Eltern wegen Gymgroß
V	V	V	V	Die Notenskala 1-5 bildet die Stärken und Schwächen (Lernbedarf) nicht ab. Die Rückmeldung ist zu wenig aussagekräftig und engt ein.
Z	V	Z	Z	Die verbale Beurteilung trifft nur selten, was sie soll, nämlich die Leistung
V	V	Z	Z	Diese Entscheidung sollte in Händen der Lehrpersonen bleiben - in

				Absprache mit den jeweiligen Eltern im Klassenforum.
V	Z	Z	Z	Eine Zahl/Note kann niemals die Leistung eines Kindes widerspiegeln. Noten dienen aber auch als Rückmeldung für die Eltern. Die erste Frage bei jeder Beurteilung ohne Noten: Welche Note ist das denn? Oft erlebt... die Eltern brauchen es...
V	V	V	V	Eine Ziffernote sagt nichts über die persönliche Leistungsentwicklung aus.
V	Z	Z	Z	Entw. des Einzelnen sollte im Vordergrund stehen-keine Klassen. Einen "Kurs" nach dem andere, erst dann in den nächsten Kurs aufsteigen, wenn der akt. pos. absolviert wird. Wir "schleppen" leider oft Kinder mit. verbal nur wenn Eltern verstehen
V	V	V	Z	Für die 4. Klasse bräuchte es beides, da es um den Wechsel in weiterführende Schulen geht. Trotzdem sollte den Kindern klar sein, warum ihre Noten so sind wie sie sind und von verbalen Beurteilungen begleitet sein.
V	V	V	V	Für eine Belohnung, weil's bequemer scheint, werden Ziffernoten gegeben. Eine aufwendigere aber sinnvollere Beurteilung ist die verbale Beurteilung.
V	V	V	V	Ich bin prinzipiell ein Gegner von Leistungsbeurteilungen...
V	V	Z	Z	Ich finde die Kombination von Lernzielkatalog und Note sehr gelungen; im ersten und am Anfang des zweiten Schuljahres könnte man gut auf die Notengebung verzichten
V	V	Z	Z	Ich finde verbale Beurteilung in der Grundstufe 1 besser da der Druck für die Kinder nicht so groß ist, aus Erfahrung haben aber viele Eltern Probleme mit verbalen Beurteilungen, da sie vieles nicht verstehen
Z	Z	Z	Z	Ich habe die Erfahrung gemacht, dass am Ende aller schriftlich verbalen Beurteilungen immer die Frage der Eltern nach der Note steht. Ergänzung durch mündliche Erläuterungen im SS- und EE-Gespräch. Besonders bei nicht-dt.spr. Eltern!
V	V	V	V	In der 4. Klasse sagen Noten für mich aus, ob ein Kind ins Gymnasium gehen kann oder nicht. Noch lieber wäre mir die Gesamtschule und dann auch keine Noten in der 4.Klasse.
V	V	V	V	ist aussagekräftiger, Teilbereiche können genauer aufgelistet werden

- V V Z Z Junge KK sind nicht in der Lage einen sinnvollen Zusammenhang zwischen ihrer erbrachten Leistung und der Note herzustellen. Sie sind alle bemüht, das Beste zu geben und sind stolz auf ihre erbrachte Leistung. Eine nicht 1 frustriert sie.
- V V Z Z KDL
- V V V V Kinder brauchen EHRliche Rückmeldungen auf ihren aktuellen Leistungsstand und ihren Fortschritt.
- V Z Z Z Kinder, die von mir verbal beurteilt wurden, fragten: und was ist das für eine Note? Bin ich gut? Ich gebe zur Note eine schriftl. Begründung bei 3 und 4
- V V V Z Leider ist zu wenig Platz um ausführlich darüber zu schreiben....!
- Z Z Z Z Leider verstehen Eltern verbale Beurteilungen nur sehr schlecht.
- - - - Meiner Meinung nach sind weder die Ziffernbenotung, noch die verbale Beurteilung eine erstrebenswerte Form der Leistungsbeurteilung im Volksschulbereich. Viel zielführender sind für mich kombinierte Varianten (zB. LFD + Noten).
- V V V V Mit einer verbalen Beurteilung kann man viel genauer aufzeigen, was das Kind schon kann und wo es noch Übungsbedarf hat.<br>Eine Note ist immer ein Kompromiss.
- V V V V Mit Verbalbeurteilung meine ich so etwas wie Pensenbuch oder Lernfortschrittsdokumentation
- V V V V Noch besser als verbale Beurteilung oder Noten, wäre die Angabe in %!  
<br>Noten sind gut, wenn Eltern kaum Deutsch sprechen.
- V V V V Noten sagen nicht, was ein Kind kann
- V V V V Noten sagen nichts aus über den tatsächlichen Leistungsstand der Schüler/innen !<br>In KDL Gesprächen können die Schüler/innen ihre Leistungen bestens zeigen ! Die Eltern werden durch die Anwesenheit beim Gespräch bestens informiert !
- V V V V Nur eine individuelle und differenzierte Benotung ist auch eine faire Benotung.
- V Z Z Z Solange sich unser gesamtes Schulsystem nicht ändert, dh sämtliche weiterführende Schulausbildungen basierend auf Stärkenförderung und Interessensförderung, halte ich die Notenbeurteilung für sinnvoller. Ich wünsche mir aber etwas anderes!

V	V	V	V	Verbale Beurteilung ist individueller und sagt mehr aus als eine Ziffernbeurteilung.
-	-	Z	Z	Verbale Beurteilung ist nur dann angemessen, wenn sie standardisiert ist. Ideal sind beide Beurteilungsformen nicht, da sie sehr subjektiv sind. Am meisten aussagekräftig sind Modelle, die entweder erreichte Lernziele/Kompetenzen abbilden.
Z	Z	Z	Z	Verbale Beurteilung schwer interpretierbar, inwiefern die Leistung gut ist und welcher Note das entsprechen würde. Viele Eltern verstehen wenig Deutsch.
V	V	V	V	Wissenschaftl. Studien belegen dass Noten weder vergleichbar+objektiv sind noch dem Fortschritt der Schülerinnen dienlich ist. Was sagt ein Einser, Zweier,... aus? Was kann ich, was kann ich nicht woran muss ich arbeiten? Soziale Kompetenz?
V	V	V	V	Wobei bei Verbalbeurteilung tatsächlich eine Form von Brief zu verstehen ist, und keinesfalls der momentan vorliegende Kompetenzraster. Noten sind auch nach der 4. Schulstufe eine viel zu vage Angabe über die Leistungen von Kindern, da ...
V	V	V	V	Würde viel lieber die Fähigkeiten der Kinder beschreiben, als ihr Unvermögen zu dokumentieren. Ziffernbeurteilung ist, wie auch die Verbalbeurteilung, nicht vergleichbar. Das Kind sollte im Mittelpunkt stehen - das vermisste ich leider in de
V	V	V	V	Ziffernnoten in der 4., da verlangt für Übertritt. Grundsätzlich ist verbal/schriftlich (LFD oder Ä.) einfach detaillierter, genauer. Außerdem ist der Druck des entsprechen Müssen und der Konkurrenzkampf eher genommen.
V	Z	Z	Z	Ziffernnoten mit einem Brief an die Kinder (Kinderzeugnis) ergänzen.
V	Z	Z	Z	Ziffernnoten von Eltern generell bevorzugt. Eltern mit Migrationshintergrund verstehen Ziffernnoten besser. Verbale Beurteilung reduziert sich letztendlich nur auf Phrasen und Satzbausteine, "Individualität" nur vorgegeben.

Tabelle 12: 1 = 1.Klasse, 2 = 2. Klasse, 3 = 3. Klasse, 4 = 4.Klasse, V = Verbalbeurteilung, Z = Ziffernnotenbeurteilung, - = „kann ich nicht beurteilen“.

## Anhang H: Begründungen für die Beurteilungspräferenzen (Laien)

Tabelle 13: Begründungen für die Angaben der Präferenzen der Leistungsbeurteilungsmethoden in den vier Volksschulklassen der Laienstichprobe (n = 30):

1	2	3	4	Begründung
Z	Z	Z	Z	Um zu vermeiden das die Kinder alle mit ihr seit eh alle super und toll beschrieben werden.
V	V	Z	Z	am besten beides
V	V	V	V	Aus pers.Erfahrung. Hatte einen sehr schlechten Pflichtschulabschluss. Konnte mich erst in Berufsschule und HTL (Abendschule) erst richtig entfalten.Ähnlich einer meiner Söhne.Dh.ein "Abstempeln" in der Volksschule halte ich für kontraprodukt
V	V	V	V	Bei einer Verbalbeurteilung kann man viel differenzierter Feedback geben und auch auf Stärken der SuS eingehen.
V	V	Z	Z	Beurteilungen befeuern meiner Meinung nach einen Vergleichswettkampf, der nur eine Wettbewerbsgesellschaft nährt. Man ist nur jemand besonderer, wenn vorgegebenen Leistungen erreicht oder übertroffen werden, ist die Botschaft.
V	-	Z	Z	da in der Arbeitswelt der Leistungsdruck immer stärker wird sollten die Kinder auch bei der Beurteilung ihrer Leistungen durch Ziffernbeurteilung darauf vorbereitet werden.
V	V	V	Z	Eine ziffer sagt für mich nicht viel aus ,durch eine verbale Beurteilung sind stärken und Schwächen erkennbar . In der 4.klasse ist vermutlich eine Ziffernbeurteilung für den Schulumstieg leichter.
V	Z	Z	Z	ergibt keinen Sinn, keine Ziffernnoten zu geben,aber jede Woche eine Ansage oder Rechentest zu machen, dies ist doch für VS Schüler ein andauernder Stress. Meine Tochter bekommt ein Zeugnis mit "Teilgenommen"- damit kann sie nichts anfangen
V	V	Z	Z	Ich denke, dass in der 1.und 2.Klasse eine verbale Beurteilung besser verstanden wird als Noten, aber diese ab der 3.Klasse durchaus vom Kind begriffen werden.

Z	Z	Z	Z	Ich denke, dass Ziffernnoten eine effizientere und klarere Leistungsbeurteilung ermöglichen..Verbalbeurteilungen mögen detailliertete Informationen über Stärken und Schwächen erlauben, sind aber einer klaren Leistungseinstufung abträglich
V	V	V	V	Ich finde eine verbale Beurteilung besser, weil man damit die Stärken und Schwächen eines Schülers viel besser ausdrücken kann als mit Ziffern.
V	V	-	-	Ich glaube eine Kombination aus Beschreibung und Ziffernnoten wäre gut. Beschreibung wegen der Information über die Fähigkeiten des Kindes, Ziffern wegen der Einordnung in das Leistungsspektrum in der Klasse, Schule, im Jahrgang
V	V	V	V	Ich kenne beide Systeme. Ich wurde mit Ziffern benotet und mein Kind verbal. Ich finde die Benotung meines Kindes viel lehrreicher, motivierender, für Kinder besser verständlich, für Eltern informativer.
V	V	V	Z	In der Volksschule können sich viele Schüler unter Noten nicht viel vorstellen. Verbale Beurteilungen sind sinnvoller. Jedoch sollte rechtzeitig auf Noten vorbereitet werden, da sich später alles um Noten dreht.
V	V	V	V	Kind soll Kind bleiben dürfen! Ich würde die positiven Fähigkeiten herausheben und einen motivierenden Bericht verfassen !geistige Nachzügler holen vieles in späteren Jahren an Leistungsfähigkeit und Wissensumfang auf
V	V	V	Z	Kinder sollten so lange wie möglich frei vom Leistungsdruck leben dürfen, es gibt genügend ander "Bewertungen" leider auch schon in der Volksschule.
V	V	Z	Z	Kombination aus Notenund verbaler Beurteilung finde ich optimal.
V	V	V	Z	Leistungsbeurteilungen nach Kriterienkatalogen oder Kompetenzraster sind aussagekräftiger als Ziffernnoten. Ziffernnoten sind nicht zwingend objektiver. Ohne Ziffernnoten 4. Kl haben KK Nachteile an der Nahtstelle
V	V	V	V	Man kann mehr beschreiben, was schon gut funktioniert und was noch mehr geübt gehört
V	V	V	V	Man kann präziser beurteilen und (wenn es richtig gemacht wird) den

				Fokus besser auf den persönlichen Leistungsvergleich legen (z.B. die Leistung im letzten Schuljahr im Vergleich zur aktuellen Leistung), statt rein auf den Vergleich mit an
V	V	V	Z	Meiner Meinung nach ist eine verbale Beurteilung immer besser wenn es die Möglichkeit gibt Rückfragen zu stellen und Talente sowie Bereiche wo Aufholbedarf besteht, besprochen werden können. Ziffernnoten 4. Klasse nötig wg öst. Schulsystem
V	V	V	V	Meiner Meinung nach kann man die "Leistungen" der SchülerInnen mit Worten viel besser beschreiben, als das mit Zifferen der Fall ist.
V	V	V	V	Meiner Meinung nach, ist die Beurteilung mittels Ziffernnoten für kleine Volksschulkinder nicht geeignet. Das würde Druck aufbauen.
V	Z	Z	Z	Noten geben meiner Meinung nach mehr klare Information.
V	V	V	V	Noten sollten immer verbal genauer beschrieben werden.
Z	Z	Z	Z	Verbale Beurteilung funktioniert vielleicht bei jungen Schülern besser. Aber dann fällt der spätere Einstieg in ein "härteres" System schwerer.
V	V	Z	Z	Verbale Beurteilung ist aussagekräftiger
V	V	V	V	Ziffernnoten sagen nichts über die Kompetenzen der SuS aus.
V	V	V	V	Ziffernnoten werden oft nach sehr unterschiedlichen Kriterien oder aufgrund unterschiedlicher Leistungsanforderungen vergeben, sind daher wenig vergleichbar
V	V	V	V	Zuviel Druck für die Kinder

Tabelle 13: 1 = 1.Klasse, 2 = 2. Klasse, 3 = 3. Klasse, 4 = 4.Klasse, V = Verbalbeurteilung, Z = Ziffernnotenbeurteilung, - = „kann ich nicht beurteilen“.

## **Anhang I: Begründungen für die Nicht-Benotung von Peters Geschichte**

Folgende unveränderten Begründungen für eine Nicht-Benotung (n = 10) wurden von den Lehrer\*innen angegeben:

- 1) „Es fehlt der Kriterienkatalog mit Punktesystem zum Beurteilen des Textes. Die Rechtschreibung kann nur unter Berücksichtigung des Grundwortschatzes erfolgen.“
- 2) „Weil ich nicht weiß, wie die Schularbeit aufbereitet wurde, gab es Lernwörter dazu, Anleitungen, außerdem brauche ich einen Referenzrahmen.“
- 3) „Weil er für mich zw. Sehr gut und Gut steht und ich letztlich mit anderen Kindern, die auch zw 1 und 2 stehen, vergleichen würde.“
- 4) „Zur Beurteilung ist das gesamte Setting notwendig. Sprachliche Fähigkeiten des Schülers, Behandlung bzw. Erarbeitung des Themas im Unterricht, Schwächen des Schülers, seit wann ist der Schüler im Land, welche Kriterien wurden zum Schreiben der Geschichte im Einzelfall vorgegeben, etc.“
- 5) „Die Beurteilung der Geschichte würde jetzt zu viel meiner Zeit beanspruchen, wir haben in Tiro eine Beurteilungstabelle, nach der Aufsätze beurteilt werden, das bringt mehr Gerechtigkeit, braucht aber auch eine genaue Analyse, unter anderem auch mit den gezählten Wörtern pro Aufsatz.“
- 6) „ich habe in meiner Schule noch nie so einen Aufsatz bekommen, würde aber wsl ein "Gut" geben, weil der Rest der Klasse nur bis 5 benotet werden kann.“
- 7) „Weil eine Schularbeit nicht nur aus einer Geschichte besteht, sondern nach den Bildungsstandards aus mehreren Teilen. Daraus ergibt sich dann die Note der Schularbeit.“
- 8) „Es ist aus dem Kontext nicht möglich! Die wievielte Geschichte schrieb Peter. ...“
- 9) „Da ich als Begleitlehrerin keine Erfahrung mit Beurteilungen von Texten gemacht habe.“
- 10) „Ich vergebe Punkte für Inhalt, Ausdruck und Rechtschreibung.“