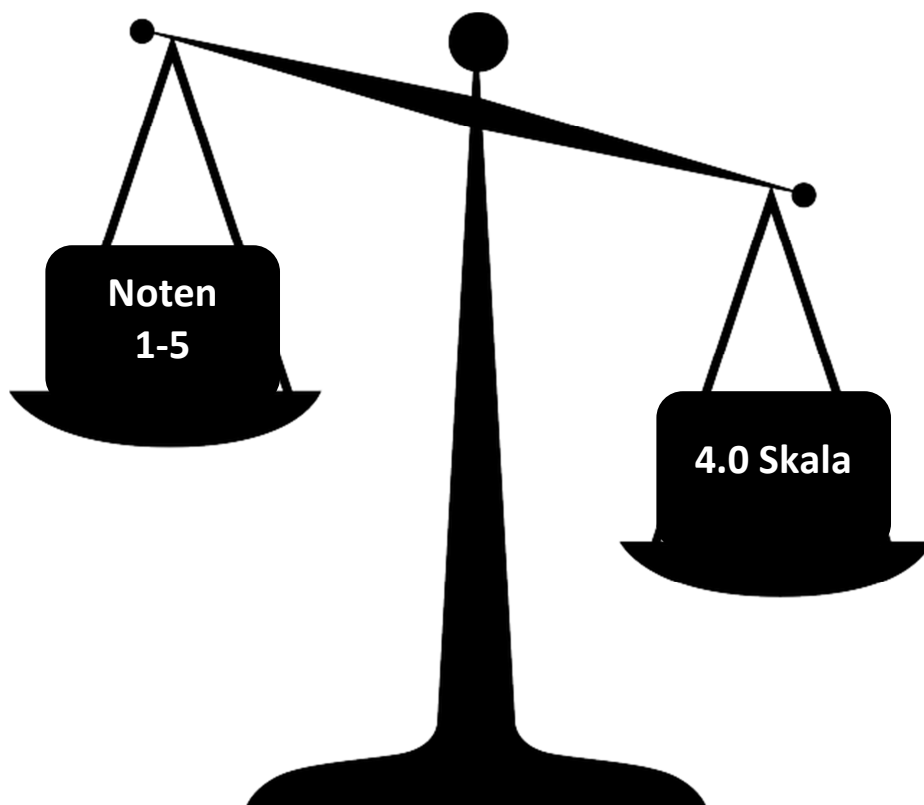


# LEISTUNGSBEURTEILUNG mit der 4.0 Skala

....

*Eine kritische Analyse  
samt Lösungsvorschlägen*



- Positionspapier -

Dr. Gerhard Ruedl

September 2015

# Inhaltsverzeichnis

VORWORT	3
GEDANKEN-BILD	4
1. Geschichten zur Leistungsbeurteilung mit der 4.0 Skala	5
<i>GESCHICHTE 1: Zwiegespräch mit einem Olympiasieger</i>	5
<i>GESCHICHTE 2: Zur Chancengleichheit bei einer Beurteilung mit der 4.0 Skala</i>	8
2. Kriterienorientierte Leistungsbeurteilung mit der 4.0 Skala nach Schlichtherle et al. (2103)	10
3. Gesetzliche Grundlagen der Leistungsbeurteilung mit der 4.0 Skala	19
3.1 Gesetzliche Grundlagen der Leistungsbeurteilung von Zeugnisnoten	23
3.2 Gesetzliche Grundlagen der Leistungsbeurteilung von Schularbeiten	24
4. Grundlagen der Ermittlung einer Gesamtnote von Schularbeiten bzw. für das Zeugnis	28
4.1 Gibt es rechtskonforme Alternativen zur Mittelwertbildung bei der Ermittlung der Zeugnisnote?	28
4.2 Zur Gewichtung der Leistungsfeststellungen	32
4.3 Entscheidungsgrundlage für die Umwandlung der 4.0 Skala in Ziffernnoten	39
4.4 Entscheidungsgrundlage für die Umwandlung der 4.0 Skala in Ziffernnoten nach grundlegenden und vertiefenden Gesichtspunkten in der 7. und 8. Schulstufe der NMS	42
5. Schlussbetrachtung	44
LITERATUR	46

## VORWORT

*„Kriterienorientierte Leistungsfeststellung“, „Kompetenzraster“, „formative statt summative Leistungsbeurteilung“, „4.0 Skala“, „Kompetenzdiagramm“ sind nur einige Begriffe, mit denen Eltern schulpflichtiger Kinder, die eine NMS besuchen, im Laufe der ersten Schulmonate möglicherweise konfrontiert werden.*

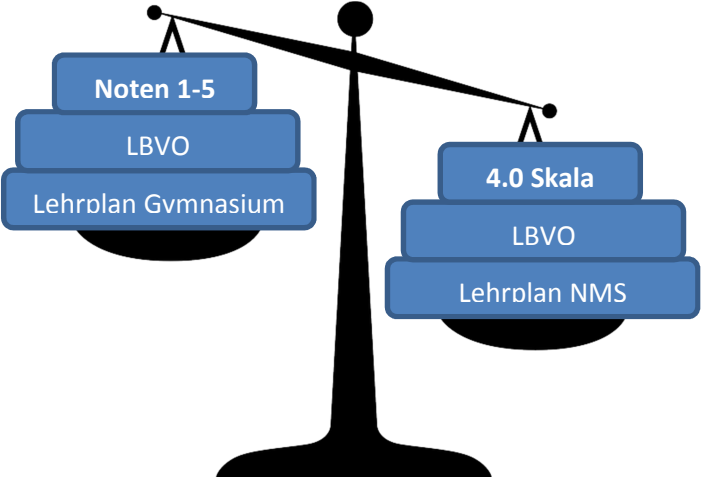
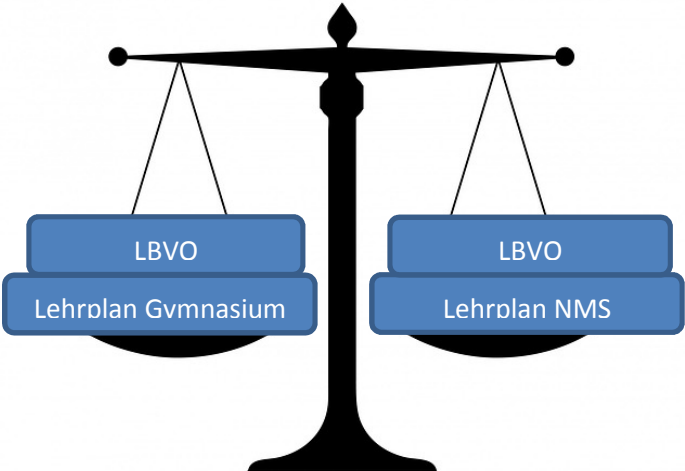
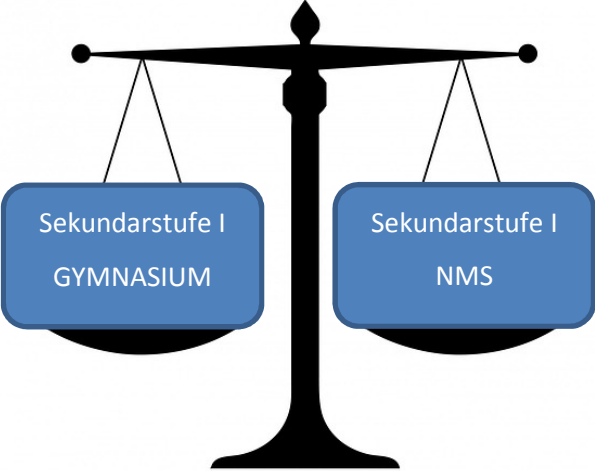
*Neben vielen positiven Aspekten eines kompetenzorientierten Unterrichtens samt kriterienorientierten Leistungsfeststellungen, welches auch einen erheblichen Mehraufwand für die Lehrpersonen bedeutet, gibt es bei der derzeit gehandhabten Leistungsbeurteilung mit der 4.0 Skala nach Ansicht des Verfassers gravierende Mängel hinsichtlich der sogenannten Beurteilungsgerechtigkeit („Gleiche Noten für gleiche Leistungen“). Genauso wie ein Arbeiter das Recht hat, bei gleicher Qualität für dieselbe Arbeit das gleiche Entgelt zu bekommen, so muss für Schüler gelten, dass diese aufgrund des einheitlichen Lehrplans für die Sekundarstufe I (Gymnasiale Unterstufe und NMS) sowie einheitlicher Bildungsstandards für vergleichbare Leistungen vergleichbare Noten bekommen, unabhängig davon, ob herkömmlich mit Noten von 1 bis 5 oder mit der 4.0 Skala beurteilt wird.*

*Das vorliegende Positionspapier ist als Diskussionsgrundlage anzusehen. Es ist das Ergebnis einer rund zwei Jahre andauernden regen Diskussion mit Lehrern, Vertretern des Landesschulrats, Schulentwicklern und Bildungsexperten.*

*Dieses Positionspapier beinhaltet neben fiktiven (?) Geschichten zum Einstieg in das Thema auch immer wieder Bezug zu den rechtlichen Grundlagen der Leistungsbeurteilungsverordnung (LBVO) und zu relevanten wissenschaftlichen Arbeiten mit dem Ziel, einerseits mögliche Probleme der derzeit gehandhabten Leistungsbeurteilung mit der 4.0 Skala aufzuzeigen sowie andererseits mögliche Lösungsvorschläge für eine „gerechte“ und vergleichbare Leistungsbeurteilung mit der 4.0 Skala vorzuschlagen.*

*Sind die in diesem Positionspapier aufgezeigten Beispiele und Kritiken nach Ansicht des Lesers zutreffend bzw. logisch nachvollziehbar und kommt es zu keinen Änderungen seitens der Schulaufsicht, so scheint es nach Ansicht des Verfassers nur eine Frage der Zeit, dass Eltern von Volksschülern ihre Schulwahl künftig nicht nur hinsichtlich des Schultyps (Gymnasium vs. Neue Mittelschule), sondern auch aufgrund der Art der Leistungsbeurteilung (herkömmlich vs. 4.0 Skala) treffen werden.*

GEDANKEN-BILD



## 1. Geschichten zur Leistungsbeurteilung mit der 4.0 Skala

### GESCHICHTE 1: Zwiegespräch mit einem Olympiasieger

*„Na Kleiner, möchtest du ein Autogramm?“*

*„Warum? Wer bist du denn?“*

*„Na, hör mal. Ich bin Ashton Eaton, der König der Leichtathleten, Olympiasieger 2012 in London und Weltrekordler im Zehnkampf.“*

*„Aha!? Und was macht so ein Zehnkämpfer?“*

*„Wir messen unsere Leistung in zehn verschiedenen Disziplinen!“*

*„Hm, also dann bin ich auch ein Zehnkämpfer!“*

*„Was? Du Zwerg willst ein Zehnkämpfer sein?“*

*„Jawohl, und nicht nur ein einfacher Zehnkämpfer wie du, sondern sogar ein mehrfacher Zehnkämpfer!“*

*„Das soll jetzt wohl ein Scherz sein?“*

*„Nein. Warum? In der Schule wird meine Leistung mehrfach in 10 Teildisziplinen ermittelt.“*

*„Was? Du machst auch Speerwerfen, Stabhochsprung und Hürdenlauf?“*

*„Nein, nein. Aber zum Beispiel im Fach Deutsch habe ich Ausdruck, Inhalt, Rechtschreibung, Lesekompetenz und andere Teildisziplinen, in denen ich bestehen muss, und die in die Wertung kommen.“*

*„Ach so, das meinst du. Kinderkram. Das wird schon nicht so schwer sein. Ich muss die besten der Welt besiegen!“*

*„Na, Moment mal. Ich muss auch meine Leistung weit über das Wesentliche hinaus bringen, damit ich erster werde bzw. einen Einser im Zeugnis bekomme!“*

*„Also ich hab‘ bei meinem Olympiasieg in London den 100-Meter-Lauf, die 400 Meter und den Weitsprung gewonnen.“*

*„Aber das sind ja nur 3 Disziplinen, in denen du der beste warst. Und da wirst du Erster?“*

*„Ja klar. Warum? Wie meinst du das jetzt?“*

*„Na hör mal, damit ich einen Einser in Deutsch bekomme, muss ich mindestens in der Hälfte der 10 Teildisziplinen eine sehr gute Leistung erbringen. Und in den restlichen Disziplinen muss ich zumindest eine gute Leistung zeigen. Nur eine durchschnittliche Leistung und ich kann den Einser vergessen!“*

*„Ehrlich? Wow! Respekt! Und ich dachte immer, die Note setzt sich quasi aus den Schularbeitsnoten und der Mitarbeit zusammen? Nur noch den Mittelwert daraus berechnen und fertig ist die Note.“*

*„Aber das ist doch Schnee von gestern! Mittelwertbildung ist out. In unserer Schule wird nun hauptsächlich der Formanstieg der Leistung für die Endnote herangezogen.“*

*„Das versteh‘ ich. Formanstieg ist gut. Ich trainiere auch das ganze Jahr hart, damit meine Form beim ersten Wettkampf passt.“*

*„Naja, das Problem ist, mein Freund geht an eine andere Schule und dort wird noch nach dem alten Schema benotet. An und für sich wäre das kein Problem, aber nächstes Jahr wollen wir beide ins Gymnasium gehen.“*

*„Ja, und?“*

*„Für dieses Gymnasium brauchen wir wahrscheinlich alles Einser. Und da hat er ganz klar einen Vorteil mit seiner Durchschnittsbenotung. Denn sobald ich in einer Teildisziplin einen 3er habe, auch wenn ich in den anderen neun Disziplinen alles Einser habe, bekomme ich im Zeugnis nur einen 2er.“*

*„Was?! Was soll denn das? Das wäre ja so, wenn ich mich für Olympia qualifizieren möchte und mein Konkurrent hätte leichtere Qualifikationslimits! Das geht doch nicht! Das ist doch ungerecht! Es müssen doch für alle die gleichen Limits gelten!“*

*„Ja, das sagt mein Vater auch!“*

*„Also ich würde Protest einlegen, wenn jemand gegen die Wettkampfstatuten verstößt.“*

*„Tja, das wollte mein Vater auch. Aber in unseren Wettkampfstatuten (Anm.: Leistungsbeurteilungsverordnung) steht nichts drinnen von der neuen Art der Benotung. Das heißt, wir können gar nicht Einspruch erheben.“*

*„Also, das verstehe ich nicht. Du hast die härteren Limits und kannst dich nicht mal gegen den Vorteil deines Konkurrenten wehren, weil nichts davon in den Wettkampfstatuten steht? Ich würde da bis zum Internationalen Sportgerichtshof gehen, um mich zu wehren!“*

*„Na klar, du bist auch ein weltberühmter Sportler! Wer bin ich dagegen?“*

*„Hör mal. Du bist du! So wie ich das jetzt sehe, hast du härter zu kämpfen als ich, um Sieger zu sein. Wärest du so nett und würdest mir ein Autogramm geben?“*

## GESCHICHTE 2: Zur Chancengleichheit bei einer Beurteilung mit der 4.0 Skala

Zwei (gleich intelligente) Kinder (ich nenne sie mal Jenny und Johnny) werden von (gleich intelligenten) Lehrern an zwei NMS nach dem gleichen Lehrplan unterrichtet (etwas zu viel des Zufalls wäre es wohl, wenn Jenny und Johnny die gleichen Lehrer hätten, die zufälligerweise an beiden Schulen dieselben Fächer unterrichten). Während bei Jenny mit der 4.0 Skala beurteilt wird, erfolgt die Beurteilung bei Johnny mit dem herkömmlichen (ich möchte noch anfügen: in der LBVO verankerten) Beurteilungssystem.

Bei beiden Kindern setzt sich jeweils die Zeugnisnote in den drei Hauptfächern aus 10 gleichberechtigten Teilnoten zusammen (bei Jenny sagt man statt Teilnoten allerdings Teilkompetenzen). In der vierten Klasse hat Jenny in jedem der drei Hauptfächer die gleichen Beurteilungen (jeweils 8 x 4.0 und 2 x 2.0) und auch Johnny hat in jedem der drei Hauptfächer die gleichen Beurteilungen (jeweils 8 x Sehr gut und 2 x Befriedigend). Jenny freut sich, denn sie hat nach dem vorgeschlagenen Umwandlungsraaster der 4.0 Skala in Ziffernnoten in jedem Hauptfach 8x einen 1er und nur 2x einer 3er bekommen. Jenny errechnet für sich einen Notendurchschnitt von 1,4 und teilt das freudig Johnny mit. Johnny grinst und gratuliert Jenny, denn er hat genau den gleichen Notendurchschnitt! Beide sprechen über die kommende gemeinsame Zeit in ihrem Wunschgymnasium! Aber oh Schreck, Jenny bekommt im Abschlusszeugnis der vierten Klasse in allen drei Hauptfächern einen 2er. Johnny hingegen bekommt in allen drei Hauptfächern einen 1er und kann somit in seinem „Wunschgymnasium“ weiter zu Schule gehen. Leider ohne Jenny!

Was ist passiert? Entsprechend der vorgeschlagenen formativen Beurteilung darf Jenny kein 2.0 haben, um ein 4.0 und somit ein „Sehr Gut“ zu erreichen. Die LehrerInnen von Jenny sind jedoch „gute“ Pädagogen. In jedem Hauptfach bekommt Jenny die sogenannte 2. Chance, um ihre Teilkompetenzen, die mit einem 2.0 beurteilt wurden, zu verbessern. In zwei Hauptfächern schafft sie jeweils ein 2.0 in ein 3.0 umzuwandeln. Das zweite 2.0 hingegen kann sie leider jeweils nicht (aus welchen Gründen auch immer) verbessern. Im dritten Hauptfach hat der Lehrer aus Zeitgründen Jenny von vornherein jedoch nur die Möglichkeit geben, eine Teilkompetenz zu verbessern, denn auch die anderen Schüler mit weniger guten Noten haben das Recht auf die 2. Chance (verständlich, denn sonst bräuchte jeder Lehrer ein halbes Semester zum Prüfen ). Jenny nützt ihre Chance im 3. Hauptfach. Das ergibt somit wiederum für jedes Hauptfach dieselben 10 Teilnoten, die sich jeweils aus 8 x 4.0, 1 x 3.0 sowie 1 x 2.0 zusammensetzen. Entsprechend der Umrechnungstabelle (4.0 = Sehr gut, 3.0 = Gut etc.) kommt Jenny auf einen Notendurchschnitt in allen drei Hauptfächern von 1,3 und ist damit „im Durchschnitt“ sogar besser als Johnny, der in sein Wunschgymnasium gehen darf, das



*Jenny leider verwehrt bleibt. Jenny versteht die Welt nicht mehr. Der Ärger und das Unverständnis bei den Eltern von Jenny ist ob der Ungleichberechtigung groß.*

*Die Geschichte von Jenny und Johnny macht im Nu die Runde bei den anderen Eltern. Die Mutter von Jimmy, der mit Jenny dieselbe Klasse besucht, meldet sich zu Wort und erzählt, dass Jimmy eigentlich auch ein sehr guter Schüler sei, aber in jedem Hauptfach einfach in einer Teilkompetenz (z.B. Inhalt in D, Geometrie in M und Sprachverständnis in E) eine deutliche Schwäche habe, die auch mit (teurer) Nachhilfe und allem Einsatz der guten Pädagogen (inklusive Teamteaching und Kultur der 2. Chance etc.) nicht zu beheben war. Jimmy hat in jedem Hauptfach in den 10 Teilkompetenzen jeweils 9 x ein 4.0 und leider jeweils 1 x ein 1.0 erreicht. Entsprechend der Umrechnungstabelle (4.0 = Sehr gut, 3.0 = Gut, 2.0 = Befriedigend, 1.0 = Genügend etc.) ergibt Jimmys Leistung einem Notendurchschnitt von 1,3 und damit ist er „im Durchschnitt“ gleich gut wie Jenny und „im Durchschnitt“ sogar besser als Johnny! Jimmys Mutter ist jedoch völlig verzweifelt. Denn Jimmy hat im Zeugnis in jedem Hauptfach einen 4er bekommen. Jimmy und Jimmys Eltern verstehen die Welt nicht mehr.*

*Was ist passiert? Entsprechend der vorgeschlagenen formativen Beurteilung darf für Noten von 1 (Sehr Gut) bis 3 (Befriedigend) keine Teilkompetenz mit weniger als einem 2.0 beurteilt werden.*

*(Fiktives und trauriges) Ende der Geschichte: Johnny geht ins Gymnasium, aber ohne Jenny macht es keinen Spaß. Mit 17 Jahren schmeißt er die Schule. Jenny möchte ins Gymnasium, wird aber nicht genommen. Sie beginnt eine Lehre als XY. Jimmy wollte eigentlich Lehrer werden, weil es ihm ein Bedürfnis war, anderen Kindern, die unter derselben Lernschwäche leiden wie er selbst, zu helfen. Mit seinem Zeugnis darf er leider nicht ins Gymnasium. Mit seinem Zeugnis bekommt er leider auch nicht den von ihm gewünschten Lehrplatz. Jimmy wird Hilfsarbeiter am Bau.*

*Fazit dieser (fiktiven?) Geschichte: Drei gleich hoch begabte Kinder mit einem gleichen (oder besseren) Notendurchschnitt werden dreimal unterschiedlich benotet, und zwar nicht nur an zwei verschiedenen Schulen, sondern sogar in derselben Klasse!*

*Fiktion oder Realität?*

## 2. Kriterienorientierte Leistungsbeurteilung mit der 4.0 Skala nach Schlichtherle et al. (2013)

In ihrem Dokument „Kriterienorientierte Leistungsfeststellung mit der 4.0-Skala“ führen die Autorinnen Birgit Schlichtherle, Veronika Weiskopf-Prantner und Tanja Westfall-Greiter (2013) einen in dieser Form in Österreich neuartigen Kompetenzraster zur Überprüfung von Leistungsfeststellungen an Neuen Mittelschulen ein:

*„Die ‚Vier-Punkt-Null-Skala‘ ist keine Erfindung der Neuen Mittelschulen, sondern ein holistischer Beurteilungsraster, der aus der Schulwirksamkeitsforschung von Marzano et al (‚what works‘) entstanden ist. Die 4.0-Skala beschreibt Leistungen anhand von sachbezogenen Kriterien auf unterschiedlichen Qualitätsstufen und bewertet diese Leistungen mit Punkten. Halbe Punktwerte geben den Hinweis darauf, dass eine Leistung im ‚Zwischenbereich‘ von zwei Qualitätsstufen liegt. Dreh- und Angelpunkt der 4.0-Skala ist die Beschreibung jener Leistung, zu der alle Schülerinnen und Schüler im Hinblick auf den Lehrplan und die Bildungsstandards befähigt werden sollen. Die Beschreibung bei 3.0 definiert dieses Zielbild. Weiters wird die Leistung definiert, die das Zielbild teilweise trifft (2.0) und die Leistung, die deutlich über dieses Zielbild hinausgeht (4.0)“ (Schlichtherle et al. 2013, S. 5).*

Die anhand eines Kompetenzrasters mit der 4.0 Skala erhobenen und bewerteten Leistungsfeststellungen werden sodann gemäß des nachfolgenden „möglichen Entscheidungsgrundlage für die 5. und 6. Schulstufe“ in Ziffernnoten umgewandelt (Schlichtherle et al. 2013, S. 18):

<b>Ergebnisse:</b>	<b>Ziffernnote:</b>
Mindestens die Hälfte der Ergebnisse ist 4.0, die restlichen Ergebnisse sind 3.0	Sehr gut
Dreiviertel der Ergebnisse sind 3.0 oder 4.0, die restlichen Ergebnisse sind nicht weniger als 2.0	Gut
Mindestens 40% der Ergebnisse sind 3.0 oder 4.0 und die restlichen 60% sind nicht weniger als 2.0	Befriedigend
Mindestens die Hälfte der Ergebnisse sind 2.0 oder höher	Genügend
	Nicht genügend

Diese „mögliche Entscheidungsgrundlage“ ist nach Ansicht des Autors zunächst in zumindest in zweierlei Hinsicht zu hinterfragen:

- 1) Es erscheint fragwürdig, dass hier von einer „möglichen Entscheidungsgrundlage“ gesprochen wird. Dadurch wird impliziert, dass auch andere Umrechnungsmodelle von Scores in Ziffernnoten möglich sind. Es stellt sich daher die Frage, wer alles diese

Umrechnungsmodelle entwickelt bzw. vorgibt? Die Bildungsexperten oder die Lerndesigner an den Schulen oder die Fachgruppe an der Schule oder jede einzelne Lehrperson etwa? Aufgrund des einheitlichen Lehrplans von gymnasialer Unterstufe und NMS sowie damit einheitlich definierter Bildungsstandards bedarf es notwendiger- und sinnvollerweise eine einheitliche bzw. zumindest vergleichbare Leistungsbeurteilung. Als Elternteil eines schulpflichtigen Kindes würde man sich auf jeden Fall eine einheitliche, gesetzlich verankerte, und zumindest von der jeweiligen zuständigen Schulaufsicht (Landesschulrat) approbiertes Umrechnungsmodell nicht nur wünschen, sondern erwarten.

- 2) Auf welchen Grundsätzen die vorgeschlagene Unterteilung der einzelnen Notenklassen basiert bzw. woher diese Unterteilung übernommen wurde, ist dem Dokument von Schlichtherle et al. (2013) nicht zu entnehmen. Unabhängig davon sollte eine Benotung auf nachvollziehbaren und vergleichbaren Notenbereiche basieren, welche einerseits in der LBVO § 14 (1) vorgegeben werden bzw. wie auch ein Beispiel von Marzano (2006, S. 122) zeigt, der nachfolgende Notenbereiche für die Umwandlung von 4.0 Scores in Noten („grades“) empfiehlt:

3.00-4.00 = A	→ <i>Notenbreite ist 1.00</i>
2.50-2.99 = B	→ <i>Notenbreite ist 0.49</i>
2.00-2.49 = C	→ <i>Notenbreite ist 0.49</i>
1.50-1.99 = D	→ <i>Notenbreite ist 0.49</i>
Below 1.50 = F	→ <i>Notenbreite ist 1.49</i>

Beim Umwandlungsmodell von Marzano (2006, S. 122), auf dessen Arbeit die 4.0 Skala basiert, fällt einerseits auf, dass der Bereich für das „Sehr gut“ (grade A) doppelt so groß wie für die weiteren Noten vom „Gut“ bis zum „Genügend“ (grades B, C, D) ist. Andererseits sind die Notenbereiche für die weiteren positiven Noten vom „Gut“ bis zum „Genügend“ (grades B, C, D) gleich groß. Im Vergleich dazu beträgt die Notenbreite bei der herkömmlichen Leistungsbeurteilung von 1 bis 5 für das Sehr gut und das Nicht genügend üblicherweise jeweils 0,5 und für die Noten Gut, Befriedigend und Genügend jeweils 1,0.

Im Folgenden wird versucht, die Notenbereiche anhand des vorgeschlagenen Umwandlungsmodells nach Schlichtherle et al. (2013) zu definieren. Zu diesem Zweck wird jeweils eine untere und obere Grenze bei 5, 6, 7, 8, 9 und 10 Teilbereichen (Kompetenzen), die in die Ziffernote einfließen, berechnet und verglichen. Die Differenz zwischen oberer und unterer Grenze ergibt die Notenbreite, die zusätzlich in Abhängigkeit der Anzahl der Teilbereiche dargestellt und verglichen wird. Aus Platzgründen werden die konkreten Berechnungen nur für das Sehr gut und das Befriedigend im Folgenden als Beispiele angeführt.

**SEHR GUT: Mindestens die Hälfte der Ergebnisse ist 4.0, die restlichen Ergebnisse sind 3.0**

**Untere Grenze:** Mindestens 50% sind 4.0, der Rest 3.0

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
	4.0,4.0,4.0,3.0,3.0	4.0,4.0,4.0,3.0,3.0, 3.0	4.0,4.0,4.0, 4.0, 3.0,3.0, 3.0	4.0,4.0,4.0, 4.0, 3.0,3.0, 3.0, 3.0	4.0,4.0,4.0, 4.0, 4.0, 3.0,3.0, 3.0, 3.0	4.0,4.0,4.0, 4.0, 4.0, 3.0,3.0, 3.0, 3.0, 3.0
Durchschnitt	3.6	3.5	3.6	3.5	3.6	3.5
Ziffernnote	Sehr gut	Sehr gut	Sehr gut	Sehr gut	Sehr gut	Sehr gut

**Obere Grenze:** 100% 4.0

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
	4.0,4.0,4.0,4.0,4.0	4.0,4.0,4.0,4.0,4.0, 4.0	4.0,4.0,4.0, 4.0, 4.0,4.0, 4.0	4.0,4.0,4.0, 4.0, 4.0,4.0, 4.0, 4.0	4.0,4.0,4.0, 4.0, 4.0, 4.0,4.0, 4.0, 4.0	4.0,4.0,4.0, 4.0, 4.0, 4.0,4.0, 4.0, 4.0, 4.0
Durchschnitt	4.0	4.0	4.0	4.0	4.0	4.0
Ziffernnote	Sehr gut	Sehr gut	Sehr gut	Sehr gut	Sehr gut	Sehr gut

**BEFRIEDIGEND: Mindestens 40% der Ergebnisse sind 3.0 oder 4.0 und die restlichen 60% Ergebnisse sind nicht weniger als 2.0**

**Untere Grenze:** 40% sind 3.0, der Rest 2.0

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
	3.0,3.0,2.0,2.0, 2.0	3.0,3.0,2.0,2.0,2.0, 2.0	3.0,3.0,3.0, 2.0, 2.0, 2.0, 2.0	3.0,3.0,3.0, 2.0, 2.0,2.0, 2.0, 2.0	3.0,3.0,3.0, 3.0, 2.0, 2.0,2.0, 2.0, 2.0	3.0,3.0,3.0, 3.0, 2.0, 2.0,2.0, 2.0, 2.0, 2.0
Durchschnitt	2.4	2.3	2.4	2.4	2.4	2.4
Ziffernnote	Befriedigend	Befriedigend	Befriedigend	Befriedigend	Befriedigend	Befriedigend

**Obere Grenze:** 40% sind 4.0, der Rest 2.0

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
	4.0,4.0,2.0,2.0, 2.0	4.0,4.0,2.0,2.0,2.0, 2.0	4.0,4.0,4.0, 2.0, 2.0, 2.0, 2.0	4.0,4.0,4.0, 2.0, 2.0,2.0, 2.0, 2.0	4.0,4.0,4.0, 4.0, 2.0, 2.0,2.0, 2.0, 2.0	4.0,4.0,4.0, 4.0, 2.0, 2.0,2.0, 2.0, 2.0, 2.0
Durchschnitt	2.8	2.7	2.9	2.8	2.9	2.8
Ziffernnote	Befriedigend	Befriedigend	Befriedigend	Befriedigend	Befriedigend	Befriedigend

**Untere und obere Grenze aller Noten in Abhängigkeit der Anzahl der Teilbereiche:**

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
Sehr gut	3.6-4.0	3.5-4.0	3.6-4.0	3.5-4.0	3.6-4.0	3.5-4.0
Gut	2.8-3.6	2.8-3.7	2.7-3.4	2.8-3.5	2.8-3.6	2.8-3.6
Befriedigend	2.4-2.8	2.3-2.7	2.4-2.9	2.4-2.8	2.4-2.9	2.4-2.8
Genügend	1.2-2.8	1.0-2.5	1.1-2.7	1.0-2.5	1.1-2.7	1.0-2.5
Nicht genügend	0.0-1.6	0.0-1.7	0.0-1.6	0.0-1.6	0.0-1.6	0.0-1.6

Ergebnis:

- 1) Die Notengrenzen hängen von der Anzahl der Teilbereiche ab, die in die Endnote einfließen;
- 2) Die jeweiligen Notenbereiche lassen sich nicht klar abgrenzen. Im Gegenteil, sie überschneiden sich teilweise deutlich;
- 3) Bei der Anzahl von 5 Teilbereichen stimmt die obere Grenze für ein Genügend mit der oberen Grenze für ein Befriedigend und mit der unteren Grenze für ein Gut überein!
- 4) Bei der Anzahl von 7 Teilbereichen besteht zwischen der oberen Grenze für ein Gut und der unteren Grenze für ein Sehr gut eine Lücke von 0.2 und die obere Grenze für eine Genügend fällt mit der unteren Grenze für ein Gut zusammen!

Fazit: Im Gegensatz zur gebräuchlichen nationalen bzw. internationalen Notengebung, in der die Leistungsbeurteilungsstufen klar voneinander abgegrenzt sind, ist das bei der Entscheidungsgrundlage von Schlichtherle et al. (2013) nicht der Fall.

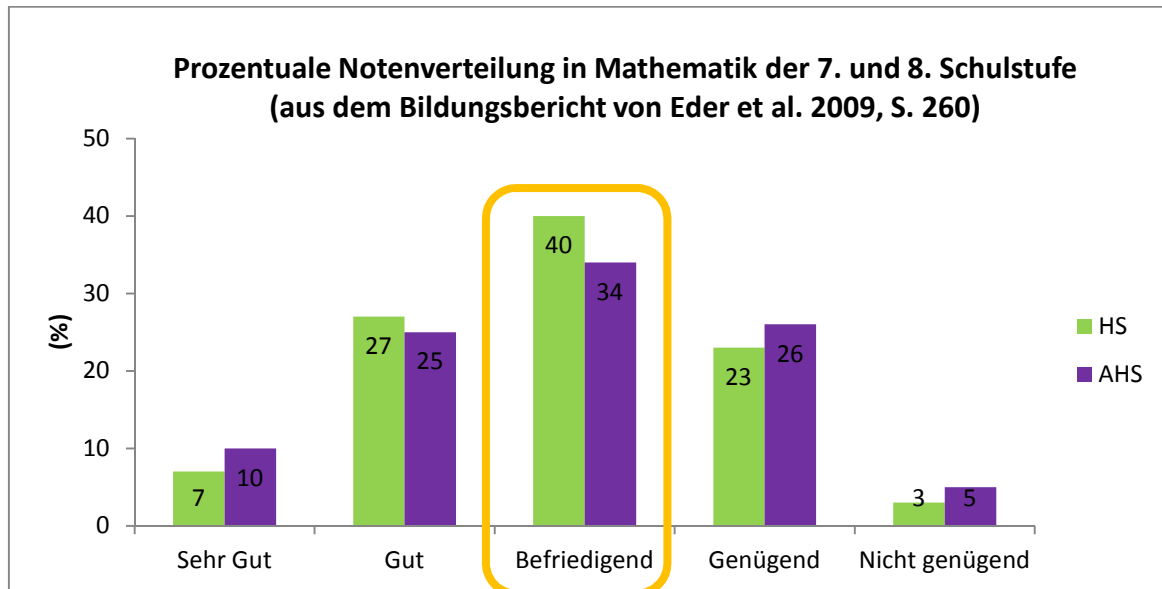
In nachfolgender Tabelle wird die Notenbreite (Differenz zwischen oberer und unterer Grenze) in Abhängigkeit der Anzahl der Teilbereiche dargestellt und verglichen.

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
<b>Sehr gut</b>	0.4	0.5	0.4	0.5	0.4	0.5
<b>Gut</b>	0.8	0.9	0.7	0.7	0.8	0.8
<b>Befriedigend</b>	0.4	0.4	0.5	0.4	0.5	0.4
<b>Genügend</b>	1.6	1.5	1.6	1.5	1.6	1.5
<b>Nicht genügend</b>	1.6	1.7	1.6	1.6	1.6	1.6

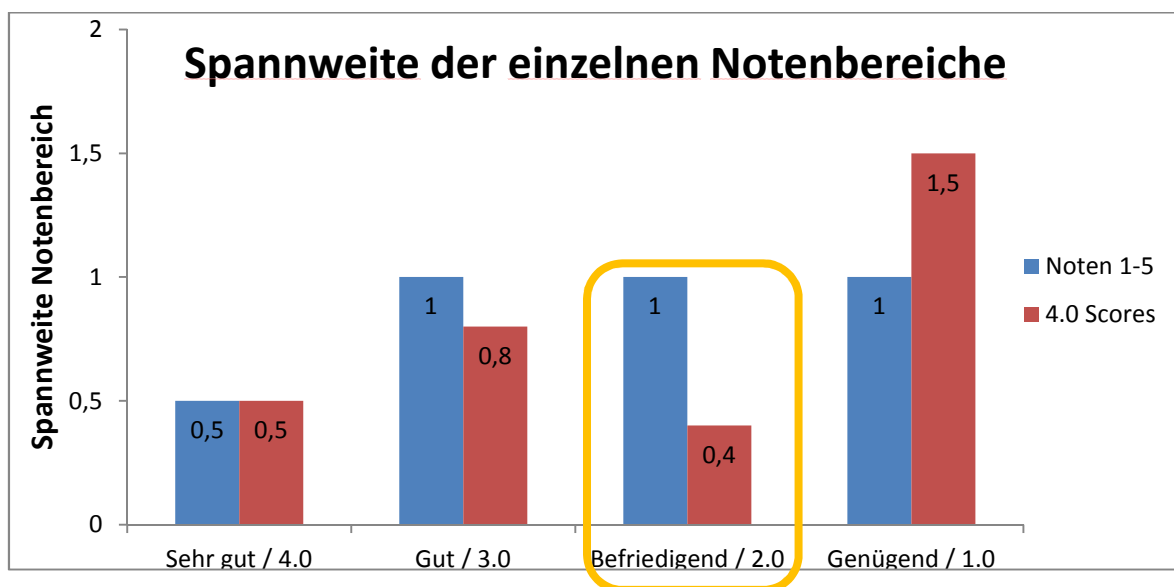
Ergebnis:

- 1) Die Spannweite einer Note ist abhängig von der Anzahl der in die Endnote einfließenden Teilnoten (Teilbereiche),
- 2) Die Spannweite für ein Gut ist teilweise doppelt so groß wie die Spannweite für ein Befriedigend (in Abhängigkeit der Teilbereiche),
- 3) Die Spannweite für ein Befriedigend ist teilweise geringer als die Spannweite für ein Sehr gut (in Abhängigkeit der Teilbereiche)
- 4) Die Spannweite für ein Genügend ist fast 4-fach größer als beim Befriedigend und fast doppelt so groß wie beim Gut.
- 5) Die Spannweite für ein Nicht genügend ist in etwa gleich groß wie für ein Genügend.

In der nachfolgenden Abbildung ist die prozentuale Notenverteilung aus dem Fach Mathematik von Gymnasiasten und Hauptschülern ersichtlich (verändert nach Eder et al. 2009, S. 260). Rund ein Drittel der Schüler liegen im Befriedigend und rund ein Viertel im Genügend. D.h. zwei Drittel aller Schüler zeigen eine durchschnittliche Leistung im Fach Mathematik.



Dieser prozentualen Notenverteilung werden in der nachfolgenden Abbildung die Notenspannweiten der herkömmlichen Benotung von 1 bis 5 sowie der 4.0 Skala (aus 10 Teilbereichen) gegenübergestellt.



Der Vergleich zeigt, dass bei der 4.0 Skala der Bereich für ein Befriedigend, in dem nachweislich (siehe Abbildung weiter oben auf dieser Seite) der größte Teil der Schüler-Leistungen im Fach

Mathematik liegt, kleiner ist als für das Sehr gut. D.h., dass „mittelmäßige“ Schüler in ihrer Leistungsbeurteilung mit der 4.0 Skala im Vergleich zu Schülern, die herkömmlich benotet werden, möglicherweise systematisch benachteiligt bzw. schlechter beurteilt werden.

Im Laufe eines Schuljahres kann es aus verschiedenen schulischen wie auch privaten Gründen vorkommen, dass Schüler bei Leistungsfeststellungen die eine oder andere „weniger gute“ Note erzielen. Wie sich eine Note, die aus dem normalen Leistungsrahmen der Kinder fällt, auswirken kann (siehe auch Geschichte 1 und 2, S. 5-9), wird nachfolgend an sogenannten potenziellen Härtefällen anhand der Entscheidungsgrundlage von Schlichtherle et al. (2013) dargestellt.

**BEISPIEL 1: Sehr gut → Definition: Mindestens die Hälfte der Ergebnisse ist 4.0, die restlichen Ergebnisse sind 3.0**

**Untere und obere Grenze des „Sehr gut“ in Abhängigkeit der Anzahl der Teilbereiche:**

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
<b>Sehr gut</b>	3.6-4.0	3.5-4.0	3.6-4.0	3.5-4.0	3.6-4.0	3.5-4.0

Annahme 1: Ein Teilbereich ist ein 2.0, alle anderen sind 4.0

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
	4.0,4.0,4.0,4.0, <b>2.0</b>	4.0,4.0,4.0,4.0,4.0, <b>2.0</b>	4.0,4.0,4.0,4.0,4.0, <b>2.0</b>	4.0,4.0,4.0,4.0,4.0, <b>2.0</b>	4.0,4.0,4.0,4.0,4.0, <b>2.0</b>	4.0,4.0,4.0,4.0,4.0, <b>2.0</b>
Durchschnitt	3.6	3.7	3.7	3.8	3.8	3.8
Ziffernnote	<b>Gut*</b>	<b>Gut*</b>	<b>Gut*</b>	<b>Gut*</b>	<b>Gut*</b>	<b>Gut*</b>

\*Für ein „Sehr Gut“ darf man gemäß Schlichtherle et al. (2013) kein 2.0 haben

Annahme 2: Ein Teilbereich ist ein 1.0, alle anderen sind 4.0

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
	4.0,4.0,4.0,4.0, <b>1.0</b>	4.0,4.0,4.0,4.0,4.0, <b>1.0</b>	4.0,4.0,4.0,4.0,4.0, <b>1.0</b>	4.0,4.0,4.0,4.0,4.0, <b>1.0</b>	4.0,4.0,4.0,4.0,4.0, <b>1.0</b>	4.0,4.0,4.0,4.0,4.0, <b>1.0</b>
Durchschnitt	3.4	3.5	3.6	3.6	3.7	3.7
Ziffernnote	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>

\*Für ein „Gut“ oder „Befriedigend“ darf man gemäß Schlichtherle et al. (2013) kein Ergebnis unter 2.0 haben

Ergebnis:

- 1) Wenn ein Teilbereich aus einem 2.0 und alle anderen Teilbereiche aus 4.0 (Annahme 1) bestehen, liegt der Durchschnitt jeweils in den Grenzen, die für das **Sehr gut** „per definitionem“ gelten. Dennoch erhält der Schüler ein **Gut!**
- 2) Wenn ein Teilbereich aus einem 1.0 und alle anderen Teilbereiche aus 4.0 (Annahme 2) bestehen, liegt der Durchschnitt (mit Ausnahme bei 5 Teilbereichen) jeweils in den Grenzen, die für das **Sehr gut** „per definitionem“ gelten. Der Schüler bekommt jedoch ein **Genügend**.

**BEISPIEL 2: Gut** → Definition: *Dreiviertel der Ergebnisse sind 3.0 oder 4.0, die restlichen Ergebnisse sind nicht weniger als 2.0*

**Untere und obere Grenze des „Gut“ in Abhängigkeit der Anzahl der Teilbereiche:**

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
<b>Gut</b>	2.8-3.6	2.8-3.7	2.7-3.4	2.8-3.5	2.8-3.6	2.8-3.6

Annahme 1: Dreiviertel sind 4.0, ein 1.0 und der Rest ist 2.0

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
	4.0,4.0,4.0,4.0, <b>1.0</b>	4.0,4.0,4.0,4.0,4.0, <b>1.0</b>	4.0,4.0,4.0, 4.0, 4.0,2.0, <b>1.0</b>	4.0,4.0,4.0, 4.0, 4.0,4.0, 2.0, <b>1.0</b>	4.0,4.0,4.0, 4.0, 4.0, 4.0,4.0, 2.0, <b>1.0</b>	4.0, 4.0,4.0, 4.0, 4.0, 4.0,4.0, 4.0, 2.0, <b>1.0</b>
Durchschnitt	3.4	3.5	3.3	3.4	3.4	3.5
Ziffernnote	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>

\*Für ein „Gut“ darf man kein Ergebnis unter 2.0 haben

Annahme 2: Dreiviertel sind 3.0, ein 1.0 und der Rest ist 2.0

	5 Teilbereiche	6 Teilbereiche	7 Teilbereiche	8 Teilbereiche	9 Teilbereiche	10 Teilbereiche
	3.0,3.0,3.0,3.0, <b>1.0</b>	3.0,3.0,3.0,3.0,3.0, <b>1.0</b>	3.0,3.0,3.0, 3.0, 3.0, 2.0, <b>1.0</b>	3.0,3.0,3.0, 3.0, 3.0,3.0, 2.0, <b>1.0</b>	3.0, 3.0, 3.0, 3.0, 3.0, 3.0,3.0, 2.0, <b>1.0</b>	3.0,3.0,3.0, 3.0, 3.0, 3.0,3.0, 3.0, 2.0, <b>1.0</b>
Durchschnitt	2.6	2.7	2.6	2.6	2.7	2.7
Ziffernnote	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>	<b>Genügend*</b>

\*Für ein „Gut“ darf man kein Ergebnis unter 2.0 haben

Ergebnis:

- 1) Wenn Dreiviertel der Teilbereiche aus 4.0 sowie aus einem 1.0 und der Rest aus 2.0 (Annahme 1) bestehen, liegt der Durchschnitt jeweils in den Grenzen, die für das **Gut** „per definitionem“ gelten. Der Schüler bekommt jedoch ein **Genügend**.
- 2) Wenn Dreiviertel der Teilbereiche aus 3.0 sowie aus einem 1.0 und der Rest aus 2.0 (Annahme 2) bestehen, liegt der Durchschnitt knapp an den Grenzen, die für das **Gut** „per definitionem“ gelten. Der Schüler bekommt jedoch ein **Genügend**.

### Zusammenfassung

- Die Entscheidungsgrundlage von Schlichtherle et al. (2013) für die Umwandlung von 4.0-Scores in eine Ziffernnote grenzt die einzelnen Notenbereiche nicht klar voneinander ab. Es kommt teilweise zu deutlichen Überschneidungen der „Notengrenzen“ und in einem Fall sogar zu einer „Lücke“.
- Die Spannweite der einzelnen Notenbereiche variiert stark und es ist keine nachvollziehbare Struktur dahinter erkennbar. Vielmehr ist die Spannweite für das Befriedigend, in welchen Bereich nachweislich ein Drittel der Schüler fällt, um das Doppelte kleiner als die Spannweite



für das Gut. Möglicherweise besteht hier die Gefahr, dass mittelmäßige („befriedigende“) Schüler entsprechend ihrer Leistung sogar unterbewertet und somit benachteiligt werden.

- Die Beispiele der „Härtefälle“ zeigen, dass ausgezeichnete und gute Leistungen nicht belohnt werden. Im Gegenteil, eine einzige mittelmäßige oder schlechte Teilbeurteilung scheint schwerer zu wiegen als die restlichen „sehr guten“ Leistungen. Sehr gute und gute Leistungen können schwächere Leistungen nicht aufwiegen. Dies könnte sich möglicherweise auf die Motivation der Schüler auswirken, ihre Leistungen zu verbessern, sobald sie in einem Teilbereich ein 1.0 bekommen.
- Die vorgeschlagene Entscheidungsgrundlage scheint zudem für die Praxis umständlich und wenig praktikabel.
- Im Vergleich zur üblichen Beurteilungspraxis an AHS (Mittelwertbildung) kommt es möglicherweise zu einer Unterbewertung der Leistungen der NMS-Schüler, da nach der Entscheidungsgrundlage nach Schlichtherle et al. (2013) ein Durchschnitt von 2.0 einem Genügend entspricht, während dessen dieses 2.0 an einer AHS mit einem Befriedigend übersetzt werden würde (siehe Schlichtherle et al., Seite 2: „Wie übersetzen wir Punkte in Noten?“).

Marzano (2006, S. 123-124.), auf dessen Arbeit die 4.0 Skala beruht, ist sich der Schwierigkeit dieses von Schlichtherle et al. (2013) vorgeschlagenen sogenannten „**Conjunctive Grading Systems**“ (siehe Tabelle unten), in dem im Vergleich zum sogenannten „**Compensatory Grading System**“ schwächere Leistungen nicht per Mittelwertberechnung kompensiert werden können, bewusst: „*Unfortunately, this system can be quite punitive in certain situations. For example, assume a student receives scores of 4.0 on all topics except one, for which she receives a score of 1.5. Using the first system, the student would receive a D. The second system is less harsh but still quite demanding. Using the second system the student would receive a C.*“

	<b>SYSTEM 1</b>		<b>SYSTEM 2</b>
<b>Grade</b>	<b>Score Pattern</b>		<b>Score Pattern</b>
<b>A</b>	No topic score below 3.0		No topic score below 3.0 and the majority 3.0 or above
<b>B</b>	No topic score below 2.5		No topic score below 2.0 and the majority 2.5 or above
<b>C</b>	No topic score below 2.0		No topic score below 1.5 and the majority 2.0 or above
<b>D</b>	No topic score below 1.5		No topic score below 1.0 and the majority 1.5 or above
<b>F</b>	Some topic score below 1.5		Some topic score below 1.0 and the majority not above 1.5

Wie die Abbildung hinsichtlich der prozentualen Verteilung in Mathematik der 7. und 8. Schulstufe (Eder et al. 2009, S. 260) auf Seite 14 zeigt, befindet sich der Großteil der Schüler (~ 60%) im Bereich des Befriedigend und des Genügend. Es ist kaum anzunehmen, dass diese Schüler nur homogene Leistungsfeststellungen in einem engbegrenzten Notenbereich im Laufe eines Schuljahres erzielen. In der Schulrealität dürften „Ausreißer“ in den Leistungsfeststellungen nach oben und unten häufig vorkommen. Die Lernmotivation von Schülern der Sekundarstufe I ist altersbedingt sicherlich zu einem Großteil auch extrinsisch von den erzielten Leistungen und Noten abhängig. Beim vorgeschlagenen System von Schlichterle et al. (2013) zur Leistungsbeurteilung wird ein Kind für eine z.B. einmalig schlechtere Leistung (z.B. 1.0) übermäßig bestraft, da es diese „schlechte“ Note nicht durch eine bessere Note beim nächsten Test ausgleichen kann. Dass dadurch die Lernmotivation nicht gesteigert wird, scheint offensichtlich. Nach Ansicht des Verfassers ist sogar das Gegenteil denkbar, dass sich dieses Kind nach einer für sein Leistungsniveau „schlechten“ Note im Bewusstsein, diese nicht mehr ausbessern zu können, möglicherweise sogar weniger anstrengen wird. Aber auch positive „Noten-Ausreißer“ berücksichtigt dieses System nicht.

An den NMS wird mit Einführung der 4.0 Skala auch die „Kultur der 2. Chance“ (siehe auch Geschichte 2, S. 8) eingeführt. D.h. Kinder können nach einer „schlechten“ Leistung diese bei einer erneuten Prüfung über denselben Stoff ausbessern. Obwohl diese „Kultur der 2. Chance“ pädagogisch sinnvoll erscheint und vermutlich auch die Härte des von Schlichterle et al. (2009) vorgeschlagenen Beurteilungssystems etwas abmildern könnte, dürfte die praktische Umsetzung aufgrund der Vielzahl der Schüler, die in jedem Fach eventuell die eine oder andere Note dann ausbessern möchten, aus zeitlichen Gründen kaum machbar sein.

Auch wenn die Leistungsbeurteilungsverordnung (LBVO) die Kompensation von weniger guten durch bessere Noten weder ausdrücklich erlaubt noch ausdrücklich verbietet (weil überhaupt nicht thematisiert), so gibt doch § 22 2g) SchUG (Schulunterrichtsgesetz) den Hinweis, dass die Kompensation einer weniger guten Note durch bessere Noten im Österreichischen Schulwesen vorgesehen ist: *„Das Jahreszeugnis hat insbesondere zu enthalten: ... die Feststellung, dass der Schüler die Schulstufe mit ausgezeichnetem Erfolg abgeschlossen hat, wenn er in mindestens der Hälfte der Pflichtgegenstände mit ‚Sehr gut‘ und in den übrigen Pflichtgegenständen mit ‚Gut‘ beurteilt wurde, wobei Beurteilungen mit ‚Befriedigend‘ diese Feststellung nicht hindern, wenn dafür gleich viele Beurteilungen mit ‚Sehr gut‘ über die Hälfte der Pflichtgegenstände hinaus vorliegen; in der Neuen Mittelschule setzt die Feststellung des ausgezeichneten Erfolges in der 7. und 8. Schulstufe eine entsprechende Beurteilung in der Vertiefung sämtlicher differenzierter Pflichtgegenstände voraus; ...“*

Ein weiterer Hinweis, dass die Kompensation von weniger guten durch bessere Noten im Bereich der NMS sogar gesetzlich vorgesehen ist, findet sich auf Seite 43 (2. Absatz) in diesem Dokument!

### 3. Gesetzliche Grundlagen der Leistungsbeurteilung mit der 4.0 Skala

Gesetzliche Grundlage: **Gemäß LBVO § 11 Abs. 2 (Fassung vom 21.01.2014) hat die Lehrperson die Leistungen der Schüler sachlich und gerecht zu beurteilen, dabei die verschiedenen fachlichen Aspekte und Beurteilungskriterien der Leistung zu berücksichtigen und so eine größtmögliche Objektivierung der Leistungsbeurteilung anzustreben.**

Neuweg (2009, S. 52) hält fest: „Für die Beurteilung von Schüler/innen/leistungen ist grundsätzlich ein **für alle Schüler/innen gleicher** lehrzielbezogener **Maßstab** anzuwenden (sog. kriteriale oder curriculare Norm) [...] Die Heranziehung der kriterialen Norm für die Bewertung von Leistungen ist in der erziehungswissenschaftlichen Literatur unumstritten. [...] Sie hilft bei der Steuerung der Lehr- und Lernprozesse, weil sie den Blick auf die Sache (...) zentriert, und hilft, die Schüler/innen zur sachlich begründeten Selbsteinschätzung hinzuführen (vgl. auch § 2 Abs. 5 LBVO). Zudem gewährleistet sie Beurteilungsgerechtigkeit: **Sie sichert gleiche Noten für gleiche Leistungen, (...).**“

Annahme: Zur Leistungsfeststellung werden lernzielorientierte Kompetenzraster in Anlehnung an Neuweg (2014, S. 14 ff.) verwendet, da entsprechend der Notendefinitionen „nicht nur wissens-, sondern vor allem kompetenzorientiert zu prüfen ist“ (Neuweg, 2014, S. 13). Ebenso sollen das Gleichwertigkeitsgebot (LBVO § 3 Abs. 5) und die Gewichtungskriterien (LBVO § 20) (siehe Neuweg 2014, S. 24 ff.) bei der Beurteilung in beiden Anekdoten von nachfolgender Geschichte 3 berücksichtigt sein. Diese beiden Annahmen sollen gewährleisten, dass die inhaltliche Qualität der Prüfungsaufgaben und deren Beurteilung in der folgenden **GESCHICHTE 3** außer Diskussion stehen.

#### Geschichte 3:

*Herr Mag. Altmann unterrichtet Mathematik in einer zweiten Klasse Gymnasium und parallel dazu in einer 2. Klasse NMS. Aufgrund des einheitlichen Lehrplans in der Sekundarstufe I muss Herr Altmann seine Stunden jeweils nur einmal vorbereiten. Zufälligerweise haben die Zwillingsschwestern Luise und Lena Herrn Altmann in Mathe. Allerdings besucht Luise nach der Scheidung und häuslichen Trennung ihrer Eltern das Gymnasium, während Lena in die NMS geht. Trotzdem lernen Luise und Lena, so oft es geht, gemeinsam auf Tests und Schularbeiten. Beide sind gleich fleißig und daher auch gleich gute Schülerinnen. Sie erzielen sogar die gleichen Noten in den Mathe-Schularbeiten. Am Ende des Schuljahrs sitzt Herr Mag. Altmann über allen im vergangenen Schuljahr beurteilten und gewichteten Leistungsfeststellungen von Luise und Lena und stellt schmunzelnd fest, dass die Zwillinge aufs Haar genau dieselben Noten „gesammelt“ haben, nämlich 6 x ein Sehr Gut und 1 x ein Befriedigend. Als Mathematiker hat Herr Altmann sofort im Kopf den Durchschnitt mit 1,3 berechnet. Während Luise*

ein Sehr Gut bekommt, stellt Mag. Altmann überrascht fest, dass er Lena aufgrund der vorgeschlagenen Entscheidungsgrundlage für die Umwandlung von 4.0 Scores in Ziffernnoten (nach Schlichtherle et al. 2013) ein Gut geben muss. An dieser Stelle muss erwähnt werden, dass an der NMS von Lena seit Beginn des Schuljahres mit der 4.0 Skala beurteilt wird.

#### Variation zu Geschichte 3:

Herr Mag. Altmann unterrichtet Mathematik in einer zweiten Klasse Gymnasium und parallel dazu in einer 2. Klasse NMS. Aufgrund des einheitlichen Lehrplans in der Sekundarstufe I muss Herr Altmann seine Stunden jeweils nur einmal vorbereiten. Zufälligerweise haben die Zwillingsschwestern Luise und Lena Herrn Altmann in Mathe. Allerdings besucht Luise nach der Scheidung und häuslichen Trennung ihrer Eltern das Gymnasium, während Lena in die NMS geht. Trotzdem lernen Luise und Lena, so oft es geht, gemeinsam auf Tests und Schularbeiten. Beide sind gleich fleißig und daher auch gleich gute Schülerinnen. Sie erzielen sogar die gleichen Noten in den Mathe-Schularbeiten, wobei beide bei der ersten Schularbeit wohl aufgrund der Trennung ihrer Eltern über ein Genügend nicht hinauskommen. Am Ende des Schuljahrs sitzt Herr Mag. Altmann über allen im vergangenen Schuljahr beurteilten und gewichteten Leistungsfeststellungen von Luise und Lena und stellt schmunzelnd fest, dass die Zwillinge aufs Haar genau dieselben Noten „gesammelt“ haben, nämlich 6 x ein Sehr Gut und 1 x ein Genügend. Als Mathematiker hat Herr Altmann sofort im Kopf den Durchschnitt mit 1,4 berechnet. Während Luise ein Sehr Gut bekommt, stellt Mag. Altmann überrascht fest, dass er Lena aufgrund der vorgeschlagenen Entscheidungsgrundlage für die Umwandlung von 4.0 Scores in Ziffernnoten (nach Schlichtherle et al. 2013) ein Genügend geben muss. An dieser Stelle muss erwähnt werden, dass an der NMS von Lena seit Beginn des Schuljahres mit der 4.0 Skala beurteilt wird.

**FAZIT 1: Die vorgeschlagene Entscheidungsgrundlage für die Umwandlung von 4.0 Scores in Ziffernnoten nach Schlichtherle et al. (2013) führt in der o.a. GESCHICHTE 3 nicht dazu, dass für gleiche Leistungen gleiche Noten im Zeugnis erzielt werden können. Eine BEURTEILUNGSGERECHTIGKEIT ist somit nicht gewährleistet und damit erscheint diese Beurteilungsform in der vorliegenden Fassung nach Ansicht des Verfassers nicht rechtskonform zu sein.**

Anmerkung: Allerdings räumt der Verfasser ein, dass möglicherweise der Vergleich von Beurteilungsformen verschiedener Schultypen trotz desselben Lehrplans mitunter schwierig ist. Um diesem Argument vorzubeugen, durfte der Mathe-Lehrer in den beiden Anekdoten an beiden Schultypen in derselben Schulstufe gleich intelligente Kinder nach demselben Lehrplan unterrichten. Kritiker könnten dennoch der Meinung sein, dass hier Äpfel mit Birnen verglichen werden. Nun,

diesem Argument wird im Folgenden entsprochen, indem zu überprüfen sein wird, ob die vorgeschlagene Entscheidungsgrundlage für die Umwandlung von 4.0 Scores in Ziffernnoten nach Schlichtherle et al. (2013) innerhalb des Klassenverbunds eine BEURTEILUNGSGERECHTIGKEIT gewährleistet.

Gesetzliche Grundlage: **Gemäß LBVO § 11 Abs. 2 hat die Lehrperson die Leistungen der Schüler sachlich und gerecht zu beurteilen, dabei die verschiedenen fachlichen Aspekte und Beurteilungskriterien der Leistung zu berücksichtigen und so eine größtmögliche Objektivierung der Leistungsbeurteilung anzustreben.**

Hypothese: Gemäß LBVO § 11, Abs. 2, ist die Leistungsbeurteilung für eine/n guten („Gut“) bzw. mittelmäßigen („Befriedigend“) oder weniger guten („Genügend“) Schüler/in gleich gerecht, sachlich und objektiv wie für eine/n sehr guten Schüler/in derselben Klasse.

Voraussetzung: Die Leistungsfeststellungen während des ganzen Schuljahres erfolgen kriterienorientiert anhand eines Kompetenzrasters mittels der 4.0 Skala. „Der Code 4.0 steht für eine Leistung, die das Zielbild übertrifft, 3.0 bedeutet, dass das Leistungszielbild getroffen und 2.0, dass das Zielbild teils getroffen wurde. 1.0 wird dann aufgezeichnet, wenn das Zielbild mit Hilfe zumindest teilweise getroffen wurde“ (Westfall-Greiter 2014, S. 3). Gleichwertigkeitsgebot (LBVO § 3 Abs. 5) und die Gewichtungskriterien (LBVO § 20) (siehe Neuweg 2014, S. 24 ff.) sind bei der Beurteilung in den nachfolgenden Beispielen berücksichtigt.

#### Beispiel A:

- ✓ Wenn Schülerinnen und Schüler (SuS) einer Klasse in allen Leistungsfeststellungen das vorab definierte Zielbild treffen und somit jeweils ein 3.0 (herkömmliche Note 2) für alle Aufgaben bekommen, so wird - wie bei der herkömmlichen Beurteilung - ein GUT vergeben.
- ✓ Wenn SuS einer Klasse in allen Leistungsfeststellungen das vorab definierte Zielbild übertreffen und somit jeweils ein 4.0 (herkömmliche Note 1) für alle Aufgaben bekommen, so wird - wie bei der herkömmlichen Beurteilung - ein SEHR GUT vergeben.
- ✓ Wenn SuS einer Klasse in allen Leistungsfeststellungen das vorab definierte Zielbild teils treffen und somit jeweils ein 2.0 (herkömmliche Note 3) für alle Aufgaben bekommen, so wäre ein BEFRIEDIGEND angebracht, auch mit Hinblick auf die Notenspannweite vom GUT zum SEHR GUT (Hinweis auf § 11,2 LBVO: gerechte und sachliche Beurteilung) sowie mit Hinblick auf die herkömmliche Benotung. Tatsächlich wird ein GENÜGEND vergeben.

- ✓ Wenn SuS einer Klasse in allen Leistungsfeststellungen das vorab definierte Zielbild mit Hilfe zumindest teilweise treffen und somit jeweils ein 1.0 (herkömmliche Note 4) für alle Aufgaben bekommen, so wäre ein Genügend angebracht, auch mit Hinblick auf die Notenspannweite vom GUT zum SEHR GUT (Hinweis auf § 11,2 LBVO: gerechte und sachliche Beurteilung) sowie mit Hinblick auf die herkömmliche Benotung. Tatsächlich wird ein NICHT GENÜGEND vergeben.

#### Beispiel B:

- ✓ Wenn SuS einer Klasse in z.B. 5 von 6 Leistungsfeststellungen das vorab definierte Zielbild übertreffen und somit jeweils ein 4.0 bekommen und in einer Leistungsfeststellung das Zielbild treffen (3.0), so wird ein SEHR GUT vergeben.
- ✓ Wenn SuS einer Klasse in z.B. 5 von 6 Leistungsfeststellungen das vorab definierte Zielbild treffen und somit jeweils ein 3.0 bekommen und in einer Leistungsfeststellung das Zielbild teils treffen (2.0), so wird ein GUT vergeben.
- ✓ Wenn SuS einer Klasse in z.B. 5 von 6 Leistungsfeststellungen das vorab definierte Zielbild teils treffen und somit jeweils ein 2.0 bekommen und in einer Leistungsfeststellung das Zielbild mit Hilfe zumindest teilweise treffen (1.0), so wäre ein BEFRIEDIGEND angebracht, auch im Vergleich zu o.a. Beispielen, die zum GUT bzw. SEHR GUT führen (Hinweis auf § 11,2 LBVO: gerechte und sachliche Beurteilung). Tatsächlich wird ein **NICHT GENÜGEND (!)** vergeben.

***FAZIT 2: Die vorgeschlagene Entscheidungsgrundlage für die Umwandlung von 4.0 Scores in Ziffernnoten nach Schlichtherle et al. (2013) gewährleistet keine BEURTEILUNGSGERECHTIGKEIT für Kinder unterschiedlichen Leistungsniveaus derselben Klasse und daher erscheint diese Beurteilungsform in der vorliegenden Fassung nach Ansicht des Verfassers nicht rechtskonform zu sein.***

**GESAMTFAZIT: Die vorgeschlagene Entscheidungsgrundlage für die Umwandlung von 4.0 Scores in Ziffernnoten nach Schlichtherle et al. (2013) gewährleistet trotz einheitlichen Lehrplans KEINE BEURTEILUNGSGERECHTIGKEIT für Kinder gleicher Leistungsniveaus zwischen verschiedenen Beurteilungsformen (Noten 1-5 bzw. 4.0 Skala) bzw. für Kinder unterschiedlichen Leistungsniveaus derselben Klasse. Daher erscheint diese Beurteilungsform in Summe in der vorliegenden Fassung nach Ansicht des Verfassers nicht rechtskonform zu sein.**

### 3.1 Gesetzliche Grundlagen der Leistungsbeurteilung von Zeugnisnoten

Laut § 14 (1) LBVO bestehen für die Beurteilung der Leistungen der Schüler folgende Beurteilungsstufen:

Sehr gut	(1),
Gut	(2),
Befriedigend	(3),
Genügend	(4),
Nicht genügend	(5).

Durch die Anführung der Notenziffer in Klammer ist nach Ansicht des Verfassers nicht nur die Zuordnung zur ausgeschriebenen Beurteilung (z.B. Sehr gut - 1) definiert, sondern auch ein konstanter Notenabstand (Notenspannweite) von jeweils einer ganzen Einheit (Ziffernwert 1) zwischen den fünf Beurteilungsstufen.

Da alternative Formen der Leistungsbeurteilung, wie z.B. die 4.0 Skala, in der LBVO derzeit nicht erwähnt werden und da für die gesamte Sekundarstufe I, sowohl für das Gymnasium als auch für Neue Mittelschulen, derselbe Lehrplan und dieselbe LBVO gelten, muss nach Ansicht des Verfassers besonders unter dem Gesichtspunkt der von Neuweg (2009, S. 52) postulierten Beurteilungsgerechtigkeit („gleiche Noten für gleiche Leistungen“) die in der LBVO definierte Anzahl der Beurteilungsstufen und deren Notenabstand auch für allfällige alternative Beurteilungsformen gelten.

Entsprechend sind die Zifferscores der 4.0 Skala in folgender Weise den in der LBVO § 14 (1) angeführten Beurteilungsstufen zuzuordnen (vgl. auch Schlichtherle et al. 2013, S. 13):

	<u>Herkömmlich</u>		<u>4.0 Skala</u>
Sehr gut	(1),	→	<b>4.0</b>
Gut	(2),	→	<b>3.0</b>
Befriedigend	(3),	→	<b>2.0</b>
Genügend	(4),	→	<b>1.0</b>
Nicht genügend	(5).	→	<b>0.0</b>

Das würde wiederum bedeuten, dass ein „Unterstufen-Gymnasiast“ oder ein NMS Schüler mit herkömmlichen Beurteilungsstufen, der in allen zur Leistungsbeurteilung herangezogenen

Leistungsfeststellungen jeweils die **Note 1** (bzw. 2, 3, 4, 5) erzielt hat, ebenso wie der NMS-Schüler, der in allen zur Leistungsbeurteilung herangezogenen Leistungsfeststellungen jeweils den **Score 4.0** (bzw. 3.0, 2.0, 1.0, 0.0) erzielt hat, im Zeugnis ein **Sehr gut** (bzw. Gut, Befriedigend, Genügend, Nicht genügend) erhält.

Unabdingbare Voraussetzung für die "Vergleichbarkeit" bzw. „Gleichberechtigung“ (LBVO § 3 Abs. 5) der einzelnen Leistungsfeststellungen ist jedenfalls deren Gewichtung gemäß LBVO §3 Abs. 5 (Berücksichtigung von Stoffumfang und Schwierigkeitsgrad) bzw. LBVO §20 (größeres Gewicht der zuletzt erbrachten Leistungen).

### 3.2 Gesetzliche Grundlagen der Leistungsbeurteilung von Schularbeiten

Gemäß LBVO § 15 (2) sind für *„die Beurteilung von schriftlichen Leistungsfeststellungen [...] nur die im § 14 Abs. 1 angeführten Beurteilungsstufen (Noten) zu verwenden und in Worten einzusetzen.“*

Das bedeutet nach Ansicht des Verfassers, dass sich auch bei Schularbeiten die Leistungsbeurteilung mit einer alternativen Beurteilungsform wie z.B. der 4.0 Skala an die in der LBVO § 14 Abs. 1 angeführten Beurteilungsstufen und die damit verbundenen Notenbreiten zu halten hat.

Insofern also bei einer schriftlichen Leistungsfeststellung (insbesondere bei Schularbeiten) einzelne Aufgaben bzw. Kompetenzen getrennt (und nicht mittels Gesamtpunkte- bzw. Gesamtfehleranzahl) beurteilt werden (z.B. Einzel-Noten bzw. Einzel-Scores für Ausdruck, Hörverständnis, Inhalt, Lesefähigkeit etc. im Fach Deutsch oder im Fach Englisch), so hat sich die finale Leistungsbeurteilung der Schularbeit anhand der erzielten Leistungsfeststellungen - nach Ansicht des Verfassers - an den in der LBVO § 14 Abs. 1 angeführten Beurteilungsstufen und den damit verbundenen Notenspannweiten zu halten.

Dem gegenüber stehen nicht evidenzbasierte „Selbstversuche“ von Lehrpersonen bzw. Schulen hinsichtlich einer eigenständigen Ermittlung der Schularbeitsnote mit der 4.0 Skala (vgl. insbesondere Westfall-Greiter 2014, S. 7 ff.), die sich weder an die in der LBVO § 14 Abs. 1 angeführten Beurteilungsstufen und die damit verbundenen Notenspannweiten halten noch eine vergleichbare fach- und schulübergreifende Benotung ermöglichen, was - nach Ansicht des Verfassers - in höchstem Maße der von Neuweg (2009, S. 52) postulierten Beurteilungsgerechtigkeit („gleiche Noten für gleiche Leistungen“) sowie der Intention der Zentralmatura (einheitliche Aufgabenstellung und einheitliche Beurteilungskriterien) widerspricht.



Im Folgenden soll diese Problematik anhand eines konkreten Schularbeiten-Beurteilungsschemas einer NMS nochmals dargestellt und diskutiert werden. Bei einer Deutsch-Schularbeit werden beispielsweise jeweils die einzelnen zu überprüfenden Kompetenzen (wie z.B. Inhalt, Ausdruck, Grammatik, Rechtschreibung, Sprachrichtigkeit von Wortarten bzw. Satzgliedern, Hörverständnis, Leseverständnis etc.) mit der 4.0 Skala beurteilt. Die gesamten erzielten Scores werden dann gemittelt und die Ziffernote wird gemäß nachfolgendem Schema ermittelt:

4,00-3,50: Sehr gut

3,49-3,00: Gut

2,99-2,50: Befriedigend

2,49-2,00: Genügend

Unter 2,00: Nicht Genügend

Bei diesem Beurteilungsschema für die Schularbeit fällt einerseits auf, dass es sich deutlich vom Beurteilungsschema für die Jahresnote nach Schlichtherle et al. (2013, siehe S. 10 in diesem Dokument) unterscheidet. Andererseits wird für das Erzielen einer positiven Note die sogenannte 50%-Regelung angewendet (50% von im Durchschnitt 4 Punkten sind 2 Punkte), für die sich laut Neuweg (2014, S. 75) in der LBVO kein Hinweis findet: *„Aus der Rechtslage ergibt sich keinerlei Anhaltspunkt für die Konvention, eine positive Note ab Erreichen der 50%-Punktemarke zu vergeben. Im Gegenteil würde diese Marke, wenn man die Verwendung von Notenschlüsseln überhaupt für zulässig hält, wohl eher unter 50% liegen müssen. Denn ein Schüler mit der Note ‚Befriedigend‘ hat nicht alle Anforderungen erfüllt, wohl aber die Anforderungen in den wesentlichen Bereichen zur Gänze. Ein Schüler mit der Note ‚Genügend‘ erfüllt diese Anforderungen überwiegend, erreicht also mehr als 50% der für ein ‚Befriedigend‘ erforderlichen Punkte.“*

In Anlehnung an Beispiel A (S. 21) und Beispiel B (S. 22) soll nachfolgend aufgezeigt werden, warum möglicherweise dieses Schularbeiten-Beurteilungsschema nicht zu einer für Kinder und Eltern nachvollziehbaren und „gerecht(fertigt)en“ Schularbeiten Benotung führt. Als Voraussetzung für die nachfolgenden Beispiele gilt, dass jede einzelne Aufgabe der Schularbeit entsprechend eines kriterienorientierten Zielbildes gestellt und beurteilt wird:

- a. Wenn Schülerinnen und Schüler (SuS) in allen gestellten Aufgaben das Zielbild erreichen und somit jeweils ein 3.0 für alle Aufgaben bekommen, so ist ein GUT angebracht. Mit einem Durchschnitt von 3.0 bekommen die SuS nach o.a. Beurteilungsschema gerade noch ein GUT für ihre Schularbeit.

- b. Wenn SuS in allen Aufgaben das Zielbild übertreffen und somit jeweils ein 4.0 für alle Aufgaben bekommen, so ist ein SEHR GUT angebracht, das die SuS auch erhalten.
- c. Wenn SuS in allen Aufgaben das Zielbild geringfügig unterschreiten und somit jeweils ein 2.0 für alle Aufgaben bekommen, so wäre ein BEFRIEDIGEND angebracht, auch mit Hinblick auf die Notenspannweite vom GUT zum SEHR GUT (Hinweis auf § 11,1 LBVO: gerechte und sachliche Beurteilung). Tatsächlich wird mit einem Durchschnitt von 2.0 nach obigem Beurteilungsschema gerade noch ein GENÜGEND vergeben.
- d. Wenn SuS z.B. in 5 von 6 Aufgaben das Zielbild erreichen und somit jeweils ein 3.0 bekommen und in einer Aufgabe das Zielbild knapp unterschreiten (2.0), so wäre ein GUT angebracht. Tatsächlich bekommen die SuS mit dem Durchschnitt von 2.8 ein BEFRIEDIGEND in der Schularbeit.
- e. Wenn SuS in 5 von 6 Aufgaben das Zielbild übertreffen und somit jeweils ein 4.0 bekommen und in einer Aufgabe das Zielbild erreichen (3.0), so ist ein SEHR GUT angebracht. Tatsächlich bekommen die SuS mit dem Durchschnitt von 3.8 auch ein SEHR GUT in der Schularbeit.
- f. Wenn SuS in 5 von 6 Aufgaben das Zielbild geringfügig unterschreiten und somit jeweils ein 2.0 bekommen und in einer Aufgabe das Zielbild deutlich unterschreiten (1.0), so wäre ein BEFRIEDIGEND angebracht. Tatsächlich bekommen die SuS mit dem Durchschnitt von 1.8 ein NICHT GENÜGEND in der Schularbeit.

Diese Beispiele sollen verdeutlichen, dass es bei diesem praktizierten Beurteilungsschema für Schularbeiten keine Beurteilungsgerechtigkeit für die SuS verschiedener Leistungsniveaus gibt. Sogar der „gute“ Schüler (Beispiel d) wird im Vergleich zum „sehr guten“ Schüler (Beispiel e) benachteiligt. Noch viel mehr betrifft diese ungleiche Beurteilung „mittelmäßige“ Schüler (Beispiel f).

Zudem besteht eine „unlogische Divergenz“ zwischen der Ziffernote in der Schularbeit und der Leistungsbeurteilung für die Jahresnote nach Schlichtherle et al. (2013). Während die SuS mit den in Beispiel d) erzielten Scores mit einem Durchschnitt von 2.8 in der Schularbeit ein Befriedigend bekommen, wird mit denselben Scores im Zeugnis ein Gut erzielt! Oder: während die SuS mit den in Beispiel f) erzielten Scores mit einem Durchschnitt von 1.8 in der Schularbeit ein Nicht genügend bekommen, wird mit denselben Scores im Zeugnis ein Genügend erzielt! Das bedeutet, dass die Ziffernote in der Schularbeit nicht den tatsächlichen Leistungsstand der SuS, der in der Leistungsbeurteilung für die Jahresnote einfließt, widerspiegelt.

Dieses o.a. Beispiel eines Schularbeiten-Beurteilungsschemas steht nach Ansicht des Verfassers im „krassen“ Gegensatz zu LBVO § 11,1 (sachlich und gerechte Beurteilung) sowie LBVO § 11, 3 (Beurteilung darf nicht entmutigen bzw. Selbstachtung beeinflussen). Wieviel Emotionen, Streit und Verbote im Elternhaus könnten verhindert werden, wenn tatsächlich „genügende oder befriedigende“ Leistungen in der Schularbeit nicht mit einem „ungerechtfertigtem“ NICHT GENÜGEND versehen werden?

### **FAZIT zu Punkt 3.1 und 3.2**

Aufgrund des einheitlichen Lehrplans in der Sekundarstufe I für Gymnasien und NMS hat sich eine alternative Leistungsbeurteilung an der derzeit gültigen Fassung der LBVO und damit insbesondere an dem Grundsatz der Beurteilungsgerechtigkeit („gleiche Noten für gleiche Leistungen“) zu orientieren.

Dieser Grundsatz hat sowohl für Kinder gleicher Leistungsniveaus zwischen verschiedenen Beurteilungsformen (Noten 1-5 bzw. 4.0 Skala) als auch besonders für Kinder unterschiedlichen Leistungsniveaus derselben Klasse, die anhand der 4.0 Skala beurteilt werden, zu gelten.

Eine rechtskonforme Leistungsbeurteilung mit der 4.0 Skala hat sich an die in der LBVO § 14 (1) vorgegebenen Beurteilungsstufen und somit auch an die damit indirekt definierten Notenspannweiten zu halten. Dies hat - nach Ansicht des Verfassers - sowohl für Zeugnisnoten als auch für die Beurteilung von Schularbeiten zu gelten.

## 4. Grundlagen der Ermittlung einer Gesamtnote von Schularbeiten bzw. für das Zeugnis

### 4.1 Gibt es rechtskonforme Alternativen zur Mittelwertbildung bei der Ermittlung der Zeugnisnote?

Laut Neuweg (2009, 2014) ist die Bildung des arithmetischen Mittels über alle Leistungsfeststellungen hinweg für die Ermittlung einer Jahresnote aufgrund der qualitativ definierten und ordinalskalierten Beurteilungsstufen in der LBVO (§ 14) nicht rechtskonform.

*Dieser Sichtweise gegenüber steht jedoch die Schulpraxis, wo - möglicherweise in Ermangelung valider Alternativen - „seit jeher“ eine Mittelwertbildung zur Ermittlung der Zeugnisnote durchgeführt wird. Wäre diese Mittelwertbildung in höchstem Maße rechtswidrig, wäre in den letzten Jahrzehnten wohl des Öfteren ein „medialer Aufschrei“ zu erwarten gewesen. Insgesamt erscheint dem Verfasser die Verwendung von Begriffen wie „rechtskonform“ bzw. „rechtswidrig“ nur im Zusammenhang mit einem Rechtsgutachten bzw. Gerichtsurteil angebracht.*

Es ist nachvollziehbar, dass die Leistungsfeststellung einer rein reproduktiven Aufgabe nicht den gleichen Stellenwert wie die Leistungsfeststellung einer komplexen Aufgabe haben kann, in der die Eigenständigkeit und die Fähigkeit zur Anwendung des Wissens und Könnens der Schüler überprüft wird (vgl. Neuweg 2009, S. 53). Entsprechend würde eine Mittelwertbildung aus diesen beiden Leistungsfeststellungen jedenfalls der von Neuweg (2009, S. 52) postulierten Beurteilungsgerechtigkeit („gleiche Noten für gleiche Leistungen“) widersprechen, ähnlich wie wenn 1 Unze Silber (→ reproduktive Aufgabe) mit 1 Unze Gold (→ komplexe Aufgabe) verglichen wird. Obwohl dieselbe Menge an Metall auf dem jeweiligen Arm der Waage liegt, ist das spezifische Gewicht von Gold jedoch aufgrund seiner höheren Dichte (→ komplexe Aufgabe) höher als das Silber (→ reproduktive Aufgabe). Welche Konsequenzen ergeben sich daraus, um die „Beurteilungswaage“ im Gleichgewicht zu halten? Entweder alle Aufgaben zur Leistungsfeststellung sind (1) rein reproduktiver Natur oder (2) rein komplexer Natur oder (3) das höhere Gewicht der komplexen Aufgaben wird bei der Leistungsbeurteilung entsprechend berücksichtigt (siehe LBVO §3 Absatz 5: Berücksichtigung von Stoffumfang und Schwierigkeitsgrad bzw. LBVO §20: größeres Gewicht der zuletzt erbrachten Leistungen) .

In einer Zusammenfassung stellt Neuweg (2014, S. 111) für die Leistungsbeurteilung für eine Schulstufe daher u.a. die Forderung: „Keine Mittelwertbildung, sondern Gewichtung nach Anzahl, Stoffumfang, Schwierigkeitsgrad und Durchführungszeitpunkt“.

Leider werden die Pädagogen mit dieser Aussage dann alleine gelassen. Wie die Gewichtung nach qualitativer Aufgabenstellung, Stoffumfang bzw. Zeitpunkt der Leistungsfeststellung zu erfolgen hat, wird offensichtlich den einzelnen Lehrpersonen überlassen. Dies muss unweigerlich zu einer uneinheitlichen Beurteilung führen und steht damit im Gegensatz zu der von Neuweg (2009, S. 52) postulierten Beurteilungsgerechtigkeit („gleiche Noten für gleiche Leistungen“), insofern keine fachspezifischen bzw. schulspezifischen bzw. schulstufenspezifischen einheitlichen Beurteilungs- und Gewichtungskriterien, wie z.B. bei der Beurteilung der Zentralmatura, vorgegeben werden.

Es stellt sich zunächst die Frage nach der Art der Gewichtung: was zählt wann um wieviel mehr?

*Eine Schularbeit hat entsprechend aufgrund des Stoffumfanges mehr Gewicht als ein schriftlicher Test. Eine Schularbeit zum Ende des Semesters hat aufgrund des Zeitpunktes mehr Gewicht als eine Schularbeit zu Beginn des Semesters. Dementsprechend müsste eine Schularbeit zum Ende des Semesters zumindest dreimal mehr Gewicht als ein schriftlicher Test zu Beginn des Semesters haben.*

Wie soll nun die Gewichtung der Leistungsfeststellungen in der Praxis erfolgen, ohne dass Mittelwerte oder Prozente berechnet werden?

In obigem Beispiel könnte sich also (vereinfacht ausgedrückt) die Beurteilung für das Jahreszeugnis aus der Mitarbeitsnote, der Halbjahresnote (nach Ansicht des Autors fraglich, ob das nicht in Widerspruch zur LBVO §20 steht?), der Note des schriftlichen Tests sowie aus einer doppelt gewerteten Note aus der 1. Schularbeit und einer dreifach gewerteten Note aus der 2. Schularbeit zusammensetzen. Das klar ersichtliche Problem dabei wäre die deutliche Überrepräsentation der Schularbeiten, welche im Gegensatz zu LBVO §3 (5) (Gleichwertigkeit der Formen der Leistungsfeststellung) stehen würde. Da Schularbeiten und schriftliche Tests in der LBVO §3 (1) c) beide unter dem Punkt „besondere schriftliche Leistungsfeststellungen“ angeführt werden, ist davon auszugehen, dass sich eine Beurteilung für diesen Bereich aus den drei erwähnten und gewichteten Leistungsfeststellungen zusammensetzt. Es stellt sich erneut die Frage, wie man ohne Berechnung von Mittelwerten bzw. Prozenten zu einer nachvollziehbaren Beurteilung eines Teilbereiches bzw. in weiterer Folge zu einer Gesamtnote kommen soll?

Sollte die Qualität der beurteilten Aufgaben vergleichbar sein bzw. sollten die Aufgaben nach Umfang und Schwere (auf welche Art auch immer) gewichtet und somit auch vergleichbar sein, so scheitert eine Mittelwertbildung gemäß Neuweg (2014) dennoch „statistisch“ an der Ordinalskalierung der Schulnoten, da hier der Median als Maß der zentralen Tendenz vorgesehen ist, und nicht der Mittelwert. Es stellt sich daher die Frage, ob der Median zur Ermittlung der Zeugnisnote herangezogen werden kann?

Laut WIKIPEDIA ist der „**Median** oder **Zentralwert** [...] ein Mittelwert für Verteilungen in der Statistik. Der Median einer Auflistung von Zahlenwerten ist der Wert, der an der mittleren Stelle steht, wenn man die Werte der Größe nach sortiert. Zum Beispiel für die Werte 4, 1, 37, 2, 1 ist die Zahl 2 der Median, nämlich die mittlere Zahl in 1, 1, 2, 4, 37. Allgemein teilt ein Median einen Datensatz, eine Stichprobe oder eine Verteilung so in zwei Hälften, dass die Werte in der einen Hälfte nicht größer als der Medianwert sind, und in der anderen nicht kleiner“.

*Beispiel 1: Für die Ermittlung der Jahresnote werden bei Schüler A die (qualitativ gleichwertigen und nach Umfang und Zeitpunkt bereits gewichteten) Noten 1, 2, 3, 4, 5 einbezogen! Median und Mittelwert betragen jeweils 3.*

*Beispiel 2: Für die Ermittlung der Jahresnote werden bei Schüler B die (qualitativ gleichwertigen und nach Umfang und Zeitpunkt bereits gewichteten) Noten 1, 4, 4, 4, 4 einbezogen! Der Median beträgt 4, während der Mittelwert 3,4 beträgt.*

Da der Median gegenüber Ausreißern robust ist, würde in Beispiel 2 die einmalig erzielte sehr gute Leistung nicht berücksichtigt. Dies wäre aus pädagogischer Sicht mehr als fragwürdig. Würde diese sehr gute Leistung zusätzlich in der letzten Schularbeit eines Jahres erzielt werden, so wäre sie auch noch nach dem Umfang und Zeitpunkt verstärkt zu gewichten und müsste dementsprechend noch mehr wiegen. Entsprechend ist der Median als statistisch korrektes Maß der zentralen Tendenz hier dem arithmetischen Mittel unterlegen und in der Schulpraxis nicht sinnvoll anzuwenden!

Zudem könnte der Median zur Notenermittlung nur bei einer ungeraden Anzahl an Leistungsfeststellungen herangezogen werden, da bei einer geraden Anzahl von „Teilnoten“ der Median aus dem Mittelwert der beiden mittleren Teilnoten gebildet wird. Nach Neuweg (2014, S. 73) ist jedoch zu beachten, „dass die Vergabe von ‚Zwischennoten‘ rechtlich nicht vorgesehen ist.“

Den Schülern muss aus pädagogischer Sichtweise die Möglichkeit gegeben werden, bei qualitativ vergleichbaren Aufgaben Schwächen in einem Teilbereich durch Stärken in einem anderen

Teilbereich ausgleichen zu können, auch mit Hinblick auf LBVO §2(5) (Leistungsfeststellungen haben zu sachlich begründeter Selbsteinschätzung der Schüler hinzuführen). Dies ist durch die Berechnung des Medians nicht möglich, wie o.a. Beispiel verdeutlicht.

***Entsprechend erscheint nach Meinung des Verfassers eine Gewichtung nach Komplexität, Umfang und Zeitpunkt der erzielten Leistungsfeststellungen z.B. mit Prozenten und nachfolgender Mittelwertbildung der Teilnoten qualitativ vergleichbarer Teilbereiche der in der Praxis gängigste, ehrlichste - weil nachvollziehbarste - Weg, der der geforderten Beurteilungsgerechtigkeit („gleiche Noten für gleiche bzw. vergleichbare Leistungen“) am ehesten entspricht!***

Nach Ansicht des Verfassers sollte weniger die Mittelwertbildung als vielmehr die Gewichtung einfacher (reproduktiver) und komplexer Aufgabenstellungen und somit deren Vergleichbarkeit im Fokus stehen und einheitlich geregelt werden. Eine Gewichtung nach Komplexität, Umfang und Zeitpunkt erscheint mittels Prozentanteilen sowie anschließender Mittelwertbildung am einfachsten und nachvollziehbarsten (unter der Voraussetzung, dass diese Gewichtung den SuS und Eltern zu Beginn des Schuljahres mitgeteilt bzw. offengelegt wurde).

Konkret könnte sich die Teilnote aus o.a. Beispiel für den Bereich „besondere schriftliche Leistungsfeststellungen“ aus 20% der Beurteilung des schriftlichen Tests zu Semesterbeginn, aus 35% der Beurteilung der 1. Schularbeit und aus 45% der Beurteilung der 2. Schularbeit zusammensetzen, womit die Gewichtung nach Umfang, Komplexität und Zeitpunkt berücksichtigt wäre.

Angenommen die Qualität aller beurteilten Aufgaben, die für die Ermittlung der Zeugnisnote zur Verfügung stehen, ist einheitlich bzw. vergleichbar, dann würde ein Schüler mit alles 1er (2er, 3er, 4er, 5er) auch ohne Mittelwertbildung ein Sehr Gut (Gut, Befriedigend, Genügend, Nicht genügend) im Zeugnis erhalten.

Auch wenn bei vergleichbarer Qualität der beurteilten Aufgaben ein Schüler in Summe alles 1er und nur einen 2er (alles 2er und einen 3er bzw. alles 3er und einen 4er bzw. alles 4er und einen 5er) am Ende des Schuljahres für die Leistungsbeurteilung zu Buche stehen hat, würde der Schüler vermutlich immer noch ohne Mittelwertbildung ein Sehr Gut (Gut, Befriedigend, Genügend) im Zeugnis erhalten.

Ab wann jedoch ist die Gesamtleistung des Schülers (bei vergleichbarer Qualität der beurteilten Aufgaben) eine Note schlechter? D.h. wo ist die Schranke vom Sehr Gut zum Gut, zum Befriedigend, zum Genügend bzw. zum Nicht Genügend? Nach Neuweg (2014, S 75) ergeben sich aus der Rechtslage „keinerlei Hinweise auf die Breite der einzelnen Notenklassen. Es ist beispielsweise keineswegs so, dass die Klasse ‚Sehr gut‘ schmaler gewählt werden müsste als die anderen positiven

Notenklassen – es sei denn, man wollte die Zahl der sehr guten Noten auf diese Weise verknappen.“ Allerdings scheint es widersinnig zu sein, dass, wie vom Verfasser auf S. 13-14 für den Beurteilungsraster mit der 4.0 Skala gezeigt wurde, die Notenbreite des Befriedigens kleiner ist als für das Sehr Gut!

Wie soll man vorgehen, um zu einer gerechtfertigten und vergleichbaren Zeugnisnote zu gelangen, wenn die SuS keine homogenen Leistungsbeurteilungen in den verschiedenen Kompetenzen eines Schulfaches aufweisen (z.B. im Fach Deutsch die Note 2 für die Lesefähigkeit, Note 4 für Inhalt, Note 3 für Grammatik, Note 1 für Rechtschreibung, Note 4 für Ausdruck etc.)?

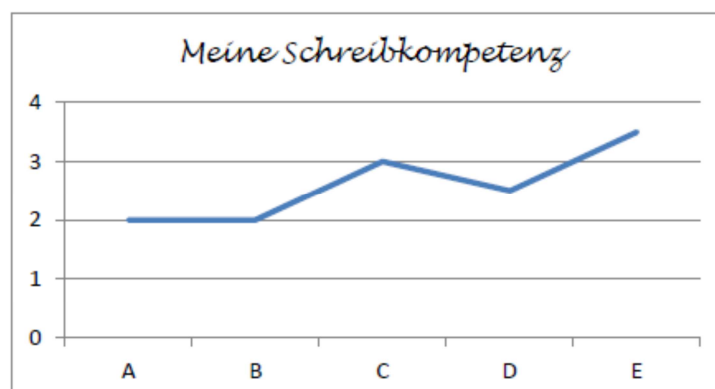
Welche praktikable, nachvollziehbare und vergleichbare (innerhalb der Klasse, innerhalb der Schulfächer und zwischen den Schultypen) Ermittlung einer Gesamtbeurteilung (gilt nicht für die Beurteilung der einzelnen Kompetenzen: siehe nächstes Kapitel) für ein Schulfach außer einer Mittelwertberechnung bleibt übrig?

**KEINE** – nach Ansicht des Verfassers!

#### 4.2 Zur Gewichtung der Leistungsfeststellungen

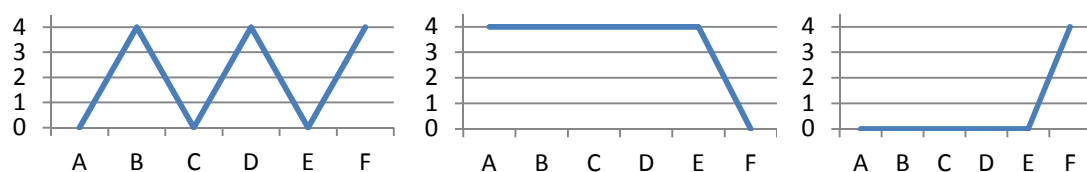
Bei der kriterienorientierten Leistungsfeststellung mit der 4.0 Skala nach Schlichtherle et al. (2013) werden die Entwicklungsverläufe verschiedener Kompetenzen eines Schulfaches (z.B. Hörverständnis, Leseverständnis, Inhalt, Ausdruck etc. im Fach Deutsch) in der Praxis so beurteilt, dass der Mittelwert aus allen erzielten Scores für einen Kompetenzbereich gebildet wird und in weiterer Folge aus den einzelnen Kompetenzen eine Gesamtnote ermittelt/gemittelt wird.

„Widersinnig“ erscheint bei der o.a. Beurteilungspraxis einer Kompetenz allerdings, dass der „formative“ Leistungsverlauf eines Kompetenzdiagramms (siehe Abbildung unten aus: Schlichtherle et al. 2013, S. 15) mittels „summativer“ Mittelwertbildung beurteilt wird, obwohl diese sogenannte „summative“ Leistungsbeurteilung von den Proponenten der 4.0 Skala regelmäßig als nicht rechtskonform bezeichnet wird (vgl. Westfall-Greiter 2013, S. 3 bzw. 9).





Eine einfache Mittelwertberechnung von 5 Leistungsfeststellungen (Zeitpunkte A-B in o.a. Abbildung) einer Kompetenz (z.B. Schreibkompetenz im Fach Deutsch) erscheint nur sinnvoll bzw. aussagekräftig, wenn die Kompetenz einen annähernd gleichbleibenden Formverlauf aufweist. Wenig sinnvoll bzw. aussagekräftig erscheint die Mittelwertbildung, wenn es um (theoretisch mögliche) extreme Formen eines Kompetenzverlaufes geht, wie in u.a. drei Abbildungen ersichtlich, da bei einer formativen Beurteilung der Entwicklungsverlauf einer Kompetenz im Fokus steht. Aus diesem Grund ist eine zuletzt erbrachte Leistung höher zu werten als frühere, da die zuletzt erbrachte Leistung auf dem Wissen früherer Inhalte aufbaut.



Um diesen Verlauf bei der Beurteilung der jeweiligen Kompetenz entsprechend zu berücksichtigen, empfiehlt Marzano (2006, S. 97f. bzw. Technical Note 5.1 auf S. 152) ein „power law of learning“, eine Potenzrechnung, wobei spätere Leistungsfeststellungen automatisch mehr gewichtet werden. Die Notwendigkeit einer solchen Berechnung im Vergleich zu einer einfachen Mittelwertberechnung ist anhand des folgenden Beispiels von 4 Schülern mit denselben Leistungsfeststellungen zu 4 Zeitpunkten (verändert nach: <http://rosehenrie.swsd.wikispaces.net/file/view/Power+Law+.pdf>) deutlich ersichtlich:

Alle 4 Schüler haben bei den 4 Leistungsfeststellungen jeweils dieselben Scores erzielt, nämlich jeweils ein 1.0, 2.0, 3.0 und 4.0, womit der Durchschnitt bei allen vier Schülern 2.5 beträgt. Berücksichtigt man jedoch, dass die 4 Schüler dieselben Scores zumeist zu einem anderen Zeitpunkt erzielt haben (siehe nachfolgende Tabelle), so bedarf es einer Berechnung mittels eines „power laws“, um eine adäquate und gerechte Beurteilung zu erzielen.

	<i>Zeitpunkt 1</i>	<i>Zeitpunkt 2</i>	<i>Zeitpunkt 3</i>	<i>Zeitpunkt 4</i>	<i>Mittelwert</i>	<i>Power Law Score</i>
<b>Schüler A</b>	1.0	2.0	3.0	4.0	2.5	<b>4.00</b>
<b>Schüler B</b>	1.0	3.0	2.0	4.0	2.5	<b>3.66</b>
<b>Schüler C</b>	2.0	4.0	1.0	3.0	2.5	<b>2.16</b>
<b>Schüler D</b>	4.0	3.0	2.0	1.0	2.5	<b>1.28</b>

Während Schüler A kontinuierlich und Schüler B zumindest größtenteils in ihren Leistungen besser wurden, erzielte Schüler D kontinuierlich geringere Scores. Dieser unterschiedliche Verlauf der Leistungen der 4 Schüler wird durch die Berechnung mittels „power law“ deutlich und nachvollziehbar berücksichtigt. Es scheint daher verwunderlich, dass diese Methode von den

Proponenten der 4.0 Skala für Österreich offensichtlich nicht „übernommen“ wurde. Generell ist nach Ansicht des Verfassers aus wissenschaftlicher Sicht zu kritisieren, dass ein sicherlich interessantes und fortschrittliches Modell aus den USA ohne entsprechende Validierungsstudien in Österreich eingeführt wird. Das dies nicht so ohne weiteres möglich erscheint, liegt nicht zuletzt auch an den Vorgaben der LBVO. Während gemäß LBVO §20 den zuletzt erbrachten Leistungen ein größeres Gewicht zukommt, was durch eine Berechnung mit einem „power law“ sehr gut berücksichtigt wird, kann die zusätzliche Gewichtung z.B. von Schularbeiten gemäß LBVO §3 Abs. 5 (Berücksichtigung von Stoffumfang und Schwierigkeitsgrad) durch eine Berechnung mittels „power law“ im Verlauf einer Kompetenz nicht ausreichend bzw. extra berücksichtigt werden. Es wird daher nachfolgend der Versuch (Vorschlag) gemacht, wie unter Berücksichtigung von LBVO §20 und LBVO §3 Abs. 5 dennoch der Verlauf einer Kompetenz möglichst rechtskonform und nachvollziehbar „berechnet“ werden kann:

Angenommen, für die Beurteilung der Kompetenz „Schreibrichtigkeit“ im Fach Deutsch stehen gesamt 8 Leistungsfeststellungen (inklusive aus 2 Schularbeiten), die im Laufe des 2. Halbjahres erzielt wurden, zur Verfügung, so bedarf es einer entsprechenden Gewichtung nach Zeitpunkt und Stoffumfang (siehe Tabelle unten). Die beiden Schularbeiten machen gesamt 40% aus, sodass für die restlichen sechs Leistungsfeststellungen (im Rahmen von Diktaten, Aufsätzen etc.) gesamt 60% (im Durchschnitt 10% je Leistungsfeststellung) überbleiben. Berücksichtigt man bei den Schularbeiten auch den Zeitpunkt gemäß LBVO §20, so könnte die 1. Schularbeit mit 15% und die 2. Schularbeit mit 25% gewichtet werden. Die restlichen 6 Leistungsfeststellungen sollten ebenso gemäß LBVO §20 gewichtet werden, um auch den Verlauf im Kompetenz-Entwicklungsdiagramm entsprechend zu berücksichtigen, da ja jeweils auf Bekanntem aufgebaut wird.

	<b>Zeitpunkt 1</b>	<b>Zeitpunkt 2</b>	<b>Zeitpunkt 3 (1.Schularbeit)</b>	<b>Zeitpunkt 4</b>	<b>Zeitpunkt 5</b>	<b>Zeitpunkt 6</b>	<b>Zeitpunkt 7 (2.Schularbeit)</b>	<b>Zeitpunkt 8</b>
<b>Score</b>	2.0	3.0	3.0	1.0	4.0	3.0	2.0	2.0
<b>Gewichtung [%]</b>	8%	9%	15%	10%	10%	11%	25%	12%

Der gewichtete Mittelwert beträgt anhand des Beispiels in Tabelle 1 für die Kompetenz „Schreibrichtigkeit“ somit 2.45! Die Gesamtnote im Zeugnis setzt sich in weiterer Folge aus allen für das jeweilige Fach relevanten gewichteten Kompetenzen sowie der Mitarbeit zusammen.

In Anlehnung an die Tabelle auf S. 33, wo mittels power law der Entwicklungsverlauf einer Kompetenz bei vier Schülern mit denselben Leistungsfeststellungen dargestellt und verglichen wurde, wird nachfolgend versucht, bei 6 Schülern mit denselben 8 Leistungsfeststellungen (1.0, 1.0, 2.0., 2.0, 3.0, 3.0, 4.0, 4.0) anhand einer prozentualen Gewichtung hinsichtlich Zeitpunkt und

Schularbeit (vergleiche Tabelle auf S. 34) einen gewichteten Mittelwert zu berechnen (Z1-8 = Zeitpunkt 1-8; S1 + S2 = Schularbeit 1 bzw. 2). Der „normale“ Durchschnitt beträgt jeweils 2.5 für alle Schüler. Schüler A zeigt einen ansteigenden Verlauf und Schüler B einen absteigenden Verlauf. Während der Verlauf bei Schüler C n-förmig ist, weist der Schüler D einen u-förmigen Verlauf auf. Schüler E hat jeweils die besten Ergebnisse und Schüler F jeweils die schlechtesten bei den Schularbeiten.

	Z1	Z2	Z3 S1	Z4	Z5	Z6	Z7 S2	Z8	
<i>Prozentuale Gewichtung</i>	8%	9%	15%	10%	10%	11%	25%	12%	
<b>Schüler</b>									<b>Endscore</b>
A	1.0	1.0	2.0	2.0	3.0	3.0	4.0	4.0	<b>2.98</b>
B	4.0	4.0	3.0	3.0	2.0	2.0	1.0	1.0	<b>2.22</b>
C	1.0	2.0	3.0	4.0	4.0	3.0	2.0	1.0	<b>2.46</b>
D	4.0	3.0	2.0	1.0	1.0	2.0	3.0	4.0	<b>2.54</b>
E	1.0	2.0	4.0	3.0	1.0	2.0	4.0	3.0	<b>2.84</b>
F	4.0	2.0	1.0	3.0	4.0	2.0	1.0	3.0	<b>2.18</b>

Ergebnisse:

1. Im Vergleich von Schüler A zu Schüler B wird ein ansteigender Verlauf der Kompetenz deutlich berücksichtigt, allerdings nicht in dem Ausmaß, wie dieser Verlauf vom power law (siehe Schüler A in Tabelle auf S. 33) bewertet werden würde.
2. Der Vergleich von Schüler C und D zeigt, dass ein zu Ende des Semesters ansteigender Verlauf zu einem höheren Durchschnittsscore führt.
3. Werden Schüler E und F verglichen, so zeigt sich deutlich der Einfluss einer höheren Gewichtung der jeweiligen Kompetenz im Rahmen der Schularbeiten.

Zusammenfassend scheint der in der o.a. Tabelle gezeigte Versuch, den Kompetenzverlauf mit Hinblick auf Zeitpunkt und Stoffumfang mittels Prozenten zu gewichten, zu einer logisch nachvollziehbaren Differenzierung des Endscores bei den 6 Schülern zu führen.

Als zweite Möglichkeit einer Gewichtung nach Zeitpunkt und Umfang ohne den Gebrauch von Prozenten wird vorgeschlagen, den erzielten Score einer Leistungsfeststellung jeweils mit dem Zeitpunkt der Leistungsfeststellung zu multiplizieren, d.h. der Score der 2. Leistungsfeststellung wird mit 2 und der Score der 8. Leistungsfeststellung entsprechend mit 8 multipliziert. Da jedoch auch noch der Stoffumfang und Zeitpunkt der Schularbeiten zu Zeitpunkt 3 bzw. zu Zeitpunkt 7 berücksichtigt werden sollten, wird vorgeschlagen, den Multiplikationsfaktor der ersten Schularbeit

zum Zeitpunkt 3 um 2 ( $3 + 2 = 5$ ) und den Multiplikationsfaktor der zweiten Schularbeit zum Zeitpunkt 7 um 3 zu erhöhen ( $7 + 3 = 10$ ). Die multiplizierten und dann summierten Scores werden schließlich durch die Summe der Multiplizierungsfaktoren (in diesem Beispiel 41) dividiert. Die nachfolgende Tabelle enthält wiederum die Scores der 6 Schüler A-F sowie den jeweiligen Multiplizierungsfaktor und den gewichteten Mittelwert.

	Z1	Z2	Z3 S1	Z4	Z5	Z6	Z7 S2	Z8	
<i>Multiplizierungs- faktor</i>	$x 1$	$x 2$	$x 5$	$x 4$	$x 5$	$x 6$	$x 10$	$x 8$	$\Sigma = 41$
<b>Schüler</b>									<b>Endscore</b>
A	1.0	1.0	2.0	2.0	3.0	3.0	4.0	4.0	<b>3.07</b>
B	4.0	4.0	3.0	3.0	2.0	2.0	1.0	1.0	<b>1.92</b>
C	1.0	2.0	3.0	4.0	4.0	3.0	2.0	1.0	<b>2.49</b>
D	4.0	3.0	2.0	1.0	1.0	2.0	3.0	4.0	<b>2.51</b>
E	1.0	2.0	4.0	3.0	1.0	2.0	4.0	3.0	<b>2.88</b>
F	4.0	2.0	1.0	3.0	4.0	2.0	1.0	3.0	<b>2.22</b>

Ergebnisse:

1. Schüler A und B werden deutlicher hinsichtlich des auf- bzw. absteigenden Verlaufs differenziert im Vergleich zum Beispiel mit der prozentualen Gewichtung (siehe Tabelle auf S. 35).
2. Der Unterschied zwischen Schüler C und D ist vorhanden, fällt allerdings im Vergleich zum Beispiel mit der prozentualen Gewichtung (siehe Tabelle auf S. 35) geringer aus.
3. Der Unterschied zwischen Schüler E und F ist fällt im Vergleich zum Beispiel mit der prozentualen Gewichtung (siehe Tabelle auf S. 35) gleich aus.

Es zeigt sich, dass beide Gewichtungsbeispiele scheinbar logisch nachvollziehbare Differenzierungen aufweisen. Allerdings zeigt sich auch, dass der Endscore der einzelnen Schüler im Vergleich der beiden Beispiele jeweils von der Art der Gewichtung abhängig ist. Dies würde stark für ein einheitliches Gewichtungssystem sprechen, welches z.B. vom jeweiligen Landesschulrat empfohlen werden könnte. Aufgrund einer zumeist unterschiedlichen Anzahl an Leistungsfeststellungen in den einzelnen Kompetenzen und somit aufgrund der einfacheren Handhabung würde das 2. Modell vom Verfasser bevorzugt werden.

Beide vorgeschlagenen Berechnungsbeispiele zeigen ebenso wie das Beispiel mit dem power law (S. 33), dass eine einfache Mittelwertbildung weder den Entwicklungsverlauf einer Kompetenz noch die

laut LBVO geforderte stärkere Gewichtung nach Zeitpunkt und Umfang entsprechend berücksichtigt und dementsprechend für die Beurteilung von Einzelkompetenzen nicht zielführend erscheint.

Marzano (2006, Fig. 7.4 auf S. 132) unterscheidet zusätzlich bei der Leistungsbeurteilung von verschiedenen Kompetenzen zwischen „Average“ und „Power law score“ insofern, dass der Mittelwert (Average) bei Leistungsfeststellungen zu einem Thema, das nicht kontinuierlich über das Schuljahr bearbeitet wird (z.B. nur 2 oder 3 Leistungsfeststellungen), zu empfehlen ist. Andererseits wird der „Potenz-Score“ empfohlen, wenn ein Thema kontinuierlich (4 und mehr Leistungsfeststellungen) über das Jahr bearbeitet wird.

Stehen also alle Leistungsbeurteilungs-Scores der einzelnen Kompetenzen für die finale Leistungsbeurteilung und deren Umwandlung in eine Ziffernote fest, so stellt sich vorab noch die Frage, ob alle Kompetenzen gleichwertig in die Endbeurteilung einfließen sollen? Oder gibt es Kompetenzen, die möglicherweise aufgrund ihrer Komplexität höher zu gewichten sind (z.B. im Fach Deutsch Leseverständnis vs. Rechtschreibung)?

Marzano (2006, S. 120-121) empfiehlt zur Lösung dieses Problems den „weighted average“ (gewichteten Mittelwert), indem er den Score einer „wichtigeren“ Kompetenz einfach mit 2 multipliziert. Die gewichteten Scores werden dann wieder summiert und durch die Anzahl der Summe der Gewichtungsfaktoren dividiert (siehe Beispiel unten, aus: Marzano 2006, S.121).

<i>Measurement Topic</i>	<i>Final Topic Score</i>	<i>Weight</i>	<i>Quality Points</i>
1. <i>Probability</i>	3.5	2	7.0
2. <i>Data Analysis ...</i>	2.5	1	2.5
3. <i>Central Tendency ...</i>	2.5	2	5.0
4. <i>Measurement</i>	1.5	1	1.5
5. <i>Problem Solving</i>	3.0	3	6.0
6. <i>Patters &amp; Functions</i>	2.0	1	2.0
7. <i>Work Completion</i>	2.5	1	2.5
8. <i>Participation</i>	2.5	1	2.5
9. <i>Behavior</i>	3.5	1	3.5
<b>Totals</b>		12	32.5

Ohne Gewichtung ergibt sich für die 9 in obiger Tabelle angeführten Kompetenzen im Fach Mathematik ein Mittelwert von 1.96; werden hingegen einzelne Kompetenzen stärker gewichtet, so liegt der Mittelwert in obigen Beispiel bei 2.71 (= 32.5/12).

Als mögliche Fehlerquelle bei der Ermittlung einer Gesamtnote aus mehreren Teilnoten soll an dieser Stelle auf das mehrmalige Runden bei Mittelwertbildungen hingewiesen werden.

Angenommen die Mittelwertbildung der vorab gewichteten Scores zeigt für Kompetenz A gesamt ein 3.4, für Kompetenz B ein 3.9 und für Kompetenz C ein 3.4, so würden sich nach Rundung für die Kompetenz A ein 3.0, für die Kompetenz B ein 4.0 sowie für die Kompetenz C ein 3.0 ergeben, die in die Leistungsbeurteilung miteinfließen. Der Mittelwert dieser 3 gerundeten Teilnoten beträgt 3.33, somit würde ein GUT im Zeugnis stehen.

Wenn hingegen die „gemittelten Werte“ der Kompetenzen A, B und C „ungerundet“ zusammengezählt werden und durch die Anzahl der Kompetenzen dividiert werden, so erhält man einen Mittelwert von 3.56, entsprechend würde im Zeugnis vermutlich ein SEHR GUT stehen.

Im Sinne einer Beurteilungsgerechtigkeit sollte es hier eine einheitliche Regelung geben, und zwar, dass nur ein einmaliges Runden des „Endmittelwertes“ zielführend sein kann, da ansonsten aufgrund eines mehrmaligen Auf- bzw. Abrundens möglicherweise positive bzw. negative Leistungen der Schüler entweder zu negativ oder zu positiv beurteilt würden, wodurch die Leistungsbeurteilung in höherem Maße verfälscht werden könnte.

Marzano (2006, S. 121) empfiehlt zumindest zum nächsten halben Score auf- oder abzurunden, da einerseits Runden auf ganze Scores zu viel an Informationsverlust einer Leistung beinhaltet, andererseits geben Halbstufen-Scores (z.B. 3.5) besser Auskunft, wie die Leistung inhaltlich in einen kriterialen Beurteilungsraster eingeordnet wird. Dies beinhaltet allerdings die Voraussetzung, dass der zu überprüfende Lernstoff im Kompetenzraster mittels Halbpunkt-Scores differenziert wird.

Basierend auf den angeführten Beispielen könnte nachfolgende Vorgehensweise zu einer adäquaten Leistungsbeurteilung von allen Leistungsfeststellungen der einzelnen Kompetenzen führen:

- ✓ Festlegung aller für das jeweilige Fach relevanten Kompetenzen, die zur finalen Leistungsbeurteilung des Schulfaches herangezogen werden, und zwar vor Schulbeginn.
- ✓ Ebenso vor Schulbeginn festlegen, ob einige der vorab ausgewählten Kompetenzen aufgrund eines höheren Komplexitätsgrades einen stärkeren Einfluss (durch eine höhere Gewichtung) auf die Endnote aufweisen sollen.
- ✓ Ebenso vor Schulbeginn die (ungefähre) Anzahl der Leistungsfeststellungen (LF) für eine Kompetenz festlegen, die zur Verfügung stehen.
- ✓ Festlegen, ob der „weighted average“ nach Marzano (2009, S. 121) bei einer Anzahl von nur 2 oder 3 LF im Semester oder bei mehreren LF ein gewichteter Entwicklungsverlauf mittels „power law“, „prozentualer oder multiplikativer“ Gewichtung sinnvoll bzw. praktikabel ist.
- ✓ Festlegen, ob entweder auf Halbpunkte oder ganze Punkte gerundet wird bzw. ob mehrmaliges Runden sinnvoll ist oder zu einem Informationsverlust führt.

### 4.3 Entscheidungsgrundlage für die Umwandlung der 4.0 Skala in Ziffernnoten

Wie bereits ausführlich dargestellt hat sich nach Ansicht des Verfassers eine „rechtskonforme“ Leistungsbeurteilung mit der 4.0 Skala an den in der LBVO § 14 (1) definierten Beurteilungstufen und deren Notenspannweiten zu orientieren, um eine aufgrund des einheitlichen Lehrplans in der Sekundarstufe I vergleichbare Leistungsbeurteilung gleicher Lerninhalte zu gewährleisten.

Das bedeutet, dass ein Schüler, der in allen Leistungsfeststellungen jeweils dieselbe Note erzielt (z.B. jeweils ein Befriedigend bzw. 2.0), im Zeugnis eben diese Note (in diesem Bsp. ein Befriedigend) erhält. Allerdings zeigt sich in der Schulpraxis, dass die zur Leistungsbeurteilung herangezogenen Leistungsfeststellungen hinsichtlich der Noten bei der Mehrzahl der SuS variieren und somit inhomogen sind. Entsprechend sind adäquate Notenbereiche zu definieren.

Wie schon früher erwähnt ergeben sich gemäß Neuweg (2014, S. 75) aus *„der Rechtsgrundlage (...) keinerlei Hinweise auf die Breite der einzelnen Notenklassen. Es ist beispielsweise keineswegs so, dass die Klasse ‚Sehr gut‘ schmaler gewählt werden müsste als die anderen positiven Notenklassen – es sei denn, man wollte die Zahl der sehr guten Noten auf diese Weise verknappen.“*

Nach Ansicht des Verfassers stellt sich hier die grundsätzliche Frage, ab wann beispielsweise (schon) ein Sehr gut vergeben werden kann/soll?

Angenommen, ein Schüler hat alles 2er (3.0 Scores) in den Leistungsfeststellungen, dann erhält dieser Schüler sicherlich ein Gut im Zeugnis. Was geschieht aber, sollte dieser Schüler in einem wesentlichen Bereich (in einer wesentlichen Kompetenz) einen 1er statt einem 2er haben? Dann hätte er/sie ja zumindest in einem wesentlichen Bereich das für ein Gut definierte „Zielbild“ übertroffen. Es stellt sich die Frage, ob damit in Anlehnung an die oben ausgeführte Aussage von Neuweg (2014, S. 75) schon ein Sehr gut gerechtfertigt ist?

Laut LBVO § 14 ist das Sehr gut so definiert, dass *„Leistungen ... in weit über das Wesentliche hinausgehendem Ausmaß erfüllt“* werden müssen, während beim Gut die *„Leistungen ... in über das Wesentliche hinausgehendem Ausmaß erfüllt“* werden müssen. Des Weiteren wird beim Sehr gut eine *„deutliche Eigenständigkeit“* verlangt, während beim Gut *„merkliche Ansätze zur Eigenständigkeit“* vorhanden sein sollten.

Es stellt sich somit die Frage, was *„weit über das Wesentliche hinaus“* bedeuten kann/soll? Wenn z.B. in allen zur Leistungsbeurteilung herangezogenen (gemäß LBVO § 3 Abs. 5 gleichwertigen)

Leistungsfeststellungen mehrheitlich (bei einer ungeraden Anzahl an Leistungsfeststellungen) oder zumindest in gleichem Ausmaß (bei einer geraden Anzahl an Leistungsfeststellungen) ein 1er erzielt würde (und der Rest 2er), so wäre der Grenzwert zwischen Sehr gut und Gut in Abhängigkeit der Anzahl der Leistungsfeststellungen der u.a. Tabelle zu entnehmen:

<i>Anzahl der Teilnoten</i>	3	4	5	6	7	8	9	10	11
Grenzwert zum Gut	1,3	1,5	1,4	1,5	1,4	1,5	1,4	1,5	1,5

Der mehrheitliche Grenzwert für das Sehr gut liegt also bei einem Mittelwert von 1,5! Diese Grenze wird zumeist im Schulalltag verwendet. Allerdings soll hier darauf aufmerksam gemacht werden, dass viele Lehrer bei einem Durchschnitt von 1,5 die Note auf einen 2er abrunden. Dies würde aber nichts anderes bedeuten, als dass der Notenbereich für ein Sehr gut nur aus einer Breite von 4 Zehntel bestehen würde und somit deutlich geringer ist als die Notenbreite der anderen Klassen! Daher wird vom Verfasser vorgeschlagen, dass das Gut erst ab einem Mittelwert von 1,55 beginnt. Alle Werte darunter wären dann dem Sehr gut zuzurechnen. Entsprechend könnten die einzelnen Notenbreiten (siehe u.a. Tabelle) folgendermaßen definiert sein:

	<i>Notenbreite herkömmliche Benotung (Spannweite)</i>	<i>Notenbreite mit der 4.0 Skala (Spannweite)</i>
Sehr gut	1,00 – 1,54 (0,54)	4.00 – 3.46 (0.54)
Gut	1,55 – 2,54 (0,99)	3.45 – 2.46 (0.99)
Befriedigend	2,55 – 3,54 (0,99)	2.45 – 1.46 (0.99)
Genügend	3,55 – 4,54 (0,99)	1.45 – 0.46 (0.99)
Nicht genügend	4,55 – 5,00 (0,45)	0.45 – 0.0 (0.45)

Es könnte natürlich auch diskutiert werden, dass „weit über das Wesentliche hinaus“ als Grenze vom Sehr gut zum Gut (an der sich dann in weiterer Folge die anderen Notenbereiche orientieren) möglicherweise nur ein Drittel 1er bei alles anderen 2er bedeuten könnte. Dementsprechend wäre die Grenze zum Gut wohl bei einem Mittelwert von rund 1,7 zu setzen. Allerdings würde sich gleichzeitig einerseits die Notenspannweite des Nicht genügend entsprechend verkleinern und andererseits würde dann ein Schüler mit zwei 5er und einem 4er (bei nur 3 Teilnoten) ein Genügend bekommen. Eine positive Note sollte nach Ansicht des Verfassers jedoch voraussetzen, dass zumindest die gleiche Anzahl an 4er und 5er bei einer geraden Anzahl an Teilnoten bzw. mehr 4er als 5er bei einer ungeraden Anzahl an Teilnoten im Laufe des Schuljahres erzielt wurden (siehe u.a. Tabelle).

<i>Anzahl der Teilnoten</i>	3	4	5	6	7	8	9	10	11
Grenzwert zum Nicht genügend	4,3	4,5	4,4	4,5	4,4	4,5	4,4	4,5	4,5



Vergleicht man nun die Entscheidungsgrundlage für die Umwandlung von 4.0 Scores in Ziffernnoten von Schlichtherle et al. (2013) z.B. für ein Sehr gut („*Mindestens die Hälfte ist ein 4.0, die restlichen sind 3.0*“), so liegt hier die Grenze zwischen Gut und Sehr gut bei 3.5 (herkömmliche Note 1,5) und wäre somit mit der herkömmlichen Benotung vergleichbar. Allerdings erlaubt diese Entscheidungsgrundlage keine schlechtere Teilnote als ein 3.0 (herkömmliche Note 2) für ein Sehr gut! Das heißt, dass – wie an anderer Stelle bereits ausführlich dargestellt wurde – bei nur einem einzigen 2.0 (herkömmliche Note 3) für einen wesentlichen Teilbereich bei gleichzeitig alles 4.0 (herkömmliche Note 1) in den anderen Teilbereichen kein Sehr gut vergeben wird. Das bedeutet, dass eine in einem oder mehreren Bereichen weniger gute Leistung nicht durch bessere Leistungen in anderen Bereichen kompensiert werden kann.

Gemäß dem Vorschlag von Schlichtherle et al. (2013) ist mit einem einzigen 1.0 (herkömmliche Note 4) in einem Teilbereich gesamt sogar nur mehr ein Genügend möglich, auch wenn in allen anderen Teilbereichen jeweils ein 4.0 (herkömmliche Note 1) erzielt werden würde. Aus pädagogischer Sichtweise erscheint dies äußerst fragwürdig, da Schüler nicht motiviert sind, ihre Leistungen zu verbessern, wenn klar ist, dass Schwächen in einem Teilbereich nicht durch Stärken in anderen Teilbereichen kompensiert werden können.

## **FAZIT**

Nach Ansicht des Verfassers muss eine Beurteilungsgerechtigkeit bei gleichem Lehrplan zwischen herkömmlicher Beurteilung und Beurteilung mit der 4.0 Skala gewährleistet sein. D.h. Noten und Notenbereiche müssen vergleichbar sein. Diese elementare Voraussetzung („*elementares Grundrecht der Schüler*“) bietet der Vorschlag von Schlichtherle et al. (2013) nicht.

Sollte die Leistungsbeurteilung mit der 4.0 Skala sich nicht an die herkömmliche Leistungsbeurteilung anpassen, so ist es nach Ansicht des Verfassers - wie schon im Vorwort angesprochen - wahrscheinlich, dass Eltern von Volksschülern ihre Schulwahl nicht nur hinsichtlich des Schultyps (Gymnasium vs. Neue Mittelschule), sondern auch aufgrund der Art der Leistungsbeurteilung (herkömmlich vs. 4.0 Skala) treffen werden.

#### 4.4 Entscheidungsgrundlage für die Umwandlung der 4.0 Skala in Ziffernnoten nach grundlegenden und vertiefenden Gesichtspunkten in der 7. und 8. Schulstufe der NMS

Nach § 14a, Abs. 1-3 LBVO hat die Beurteilung in den NMS in der 7. und 8. Schulstufe nach grundlegenden und vertiefenden Gesichtspunkten zu erfolgen. D.h., wer nicht zumindest ein in der 5. und 6. Schulstufe noch „herkömmliches“ Genügend erreicht, kann nur mehr grundlegend mit Befriedigend, Genügend bzw. Nicht genügend beurteilt werden.

Überspitzt formuliert bedeutet das, dass der Gesetzgeber ein in der 5. und 6. Schulstufe noch „herkömmliches“ Nicht genügend nun weiter unterteilt in ein „**befriedigendes Nicht genügend**“ bzw. in ein „**genügendes Nicht genügend**“ bzw. in ein „**nicht genügendes Nicht genügend**“!

Die Absicht dahinter scheint offensichtlich: durch die weitere Differenzierung der Notenbereiche in „vertiefend“ und „grundlegend“ wird die Anzahl an Nicht genügend in den Zeugnissen der 8. Schulstufe NMS vermutlich drastisch reduziert werden.

Auch in diesem Bereich scheint eine fächer- und schulübergreifende Leistungsbeurteilung dringend notwendig, um eine Beurteilungsgerechtigkeit aufgrund des einheitlichen Lehrplans zwischen Gymnasiasten einerseits und NMS-Schülern andererseits (mit herkömmlicher Beurteilung sowie mit Beurteilung anhand der 4.0 Skala) zu gewährleisten. Insbesondere, da im SchOG § 21b (2) festgelegt ist, dass im „*Lehrplan (...) für die 7. und 8. Schulstufe in den differenzierten Pflichtgegenständen eine Unterscheidung nach grundlegender und vertiefter Allgemeinbildung vorzusehen (ist). Die Anforderungen der Vertiefung haben jenen der Unterstufe der allgemein bildenden höheren Schule zu entsprechen.*“

Nach Ansicht des Verfassers gibt es daher basierend auf den bisherigen Ausführungen in diesem Positionspapier zwei Möglichkeiten die Grenze zwischen vertiefend und genügend zu definieren:

- 1) Die in der Tabelle auf Seite 40 vorgesehene Notenspannweite für ein Genügend in der 5. und 6. Schulstufe gilt auch für ein **Genügend Vertiefend**. Ab einem Durchschnitt von 4,55 wird mit Grundlegend beurteilt. D.h., die grundlegende Beurteilung beginnt erst nach dem in der 5. und 6. Schulstufe noch herkömmlichen Bereich für das Genügend.
- 2) Die untere Grenze für ein **Genügend Vertiefend** ist erreicht, wenn in jedem Teilbereich, der in die Leistungsbeurteilung einfließt, zumindest ein 4er erzielt wurde bzw. wenn der Notendurchschnitt 4,0 nicht unterschritten wird! D.h., der Notenbereich für ein in der 5. und 6. Schulstufe noch herkömmliches Genügend wird nun differenziert in vertiefend und grundlegend.

Entsprechend der 2. Möglichkeit und unter Bedacht möglichst nachvollziehbarer bzw. vergleichbarer Notenbreiten in der grundlegenden Leistungsbeurteilung wird folgender Vorschlag zu Diskussion gestellt (siehe Tabelle unten):

Notenbreite: Herkömmliche Beurteilung	Notenbreite: Beurteilung mit der 4.0 Skala	Ziffernnote 7. und 8. Schulstufe
1,00 – 1,54 (0,54)	4.00 – 3.46 (0.54)	Sehr gut <b>VERTIEFEND</b>
1,55 – 2,54 (0,99)	3.45 – 2.46 (0.99)	Gut <b>VERTIEFEND</b>
2,55 – 3,54 (0,99)	2.45 – 1.46 (0.99)	Befriedigend <b>VERTIEFEND</b>
3,55 – 4,00 (0,45)	1.45 – 1.00 (0.45)	Genügend <b>VERTIEFEND</b>
4,01 – 4,34 (0,33)	0.99 – 0.66 (0.33)	Befriedigend <b>GRUNDLEGENDE</b>
4,35 – 4,67 (0,32)	0.65 – 0.33 (0.32)	Genügend <b>GRUNDLEGENDE</b>
4,68 – 5,00 (0,32)	0.32 – 0.0 (0.32)	Nicht genügend <b>GRUNDLEGENDE</b>

Wie erfolgt die Gesamtbeurteilung eines Schülers, wenn die Leistungsfeststellungen (inklusive Schularbeiten) sowohl im grundlegenden Bereich als auch im vertiefenden Bereich erfolgt sind?

Nach dem der überprüfte Lerninhalt ja immer vertieft angeboten wird und es sich bei der vertiefenden bzw. grundlegenden Beurteilung einfach um Leistungsabstufungen derselben Leistungsfeststellung handelt, wird wiederum unter dem Gesichtspunkt der Beurteilungsgerechtigkeit eine Mittelwertbildung am praktikabelsten und nachvollziehbarsten sein, weil u.a. weniger gute Ergebnisse durch bessere Ergebnisse kompensiert werden können. D.h. nach Ansicht des Verfassers nichts anderes, als dass der Gesetzgeber durch die Unterteilung der Beurteilung in der 7. und 8. Schulstufe an den NMS in Grundlegend und Vertiefend eine Mittelwertberechnung von Leistungsfeststellungen, die sowohl vertiefend bzw. grundlegend beurteilt werden, im Sinne eines „Kompensationsmechanismus“ explizit voraussetzt. Ansonsten müsste eine einzige grundlegende Leistungsbeurteilung dazu führen, dass der Schüler nur noch in Grundlegend beurteilt werden könnte.

## FAZIT

Eine Beurteilung nach grundlegenden und vertiefenden Gesichtspunkten (herkömmlich oder mit der 4.0 Skala) in den NMS hat sich in erster Linie aufgrund des einheitlichen Lehrplanes für die NMS und die gymnasiale Unterstufe an der Beurteilungsgerechtigkeit („gleiche Noten für gleiche Leistungen“) zu orientieren. D.h. gemäß SchOG § 21b (2), dass „sehr gute bis genügende“ Leistungen von SuS in der gymnasialen Unterstufe entsprechend „sehr guten bis genügenden“ Leistungen unter einem vertiefenden Gesichtspunkt von SuS in NMS gleichgesetzt werden müssen.

Für die Unterteilung der Notenbereiche in vertiefend und grundlegend ist wiederum unter dem Gesichtspunkt der Vergleichbarkeit und somit der Beurteilungsgerechtigkeit zwischen den NMS ein einheitlicher (möglichst von der Schulaufsicht vorgegebener) und somit schulübergreifender Beurteilungsschlüssel erforderlich.

## 5. Schlussbetrachtung

Kompetenzorientierter Unterricht und kriterienorientierte Leistungsfeststellungen unter Einbeziehung von Kompetenzdiagrammen scheint ein innovativer und pädagogisch sinnvoller sowie teilweise evidenzbasierter Ansatz zu sein, den Lernertrag von Schülern zu steigern. Zusätzlich hängt damit die Leistungsbeurteilung für eine Schulstufe nicht mehr hauptsächlich von den punktuellen Leistungen im Rahmen der Schularbeiten (Schularbeitsnoten) und eventuell der Mitarbeit als „Zünglein an der Waage“ ab. Vielmehr steht der Entwicklungsverlauf der für das jeweilige Schulfach relevanten Kompetenzen im Mittelpunkt, in den auch die punktuellen Leistungen dieser Kompetenz im Rahmen der Schularbeiten einfließen.

Eine immer wiederkehrende „Basisbegründung“ für die Einführung der Leistungsbeurteilung mit der 4.0 Skala scheint gemäß Westfall-Greiter (2013) der Hinweis zu sein, dass eine Mittelwertberechnung für die Notengebung nicht rechtskonform sei. Dabei wird offensichtlich übersehen, dass damit in erster Linie eine Mittelwertbildung von z.B. reproduktiven und komplexen Aufgabenstellungen ohne eine entsprechende Gewichtung gemeint ist. Es scheint nachvollziehbar, dass eine gemittelte Zeugnisnote, herkömmlich zumeist im Wesentlichen aus zwei Schularbeitsnoten, der Mitarbeit sowie der Halbjahresnote bestehend, weniger Aussagekraft über die tatsächliche(n) Kompetenz(en) in dem jeweiligen Fach bietet. Als Alternative wurde daher im Vorschlag von Schlichtherle et al. (2013) auf eine Mittelwertberechnung verzichtet. Allerdings zeigen die Beispiele und Analysen im vorliegenden Positionspapier nach Ansicht des Verfassers deutlich, dass diese „mögliche Entscheidungsgrundlage“ nach Schlichtherle et al. (2013) weder eine Beurteilungsgerechtigkeit (gleiche Noten für gleiche Leistungen) für Kinder gleicher Leistungsniveaus zwischen verschiedenen Beurteilungsformen (Noten 1-5 bzw. 4.0 Skala) noch für Kinder unterschiedlichen Leistungsniveaus derselben Klasse gewährleistet. Vereinzelt „schwächere Leistungen“ der Kinder fallen in diesem „Conjunctive Grading System“ (siehe S. 17-18 in diesem Positionspapier) viel stärker ins Gewicht als vereinzelt „bessere Leistungen“, was sich nachteilig auf die Lernmotivation der Schüler und Schülerinnen auswirken könnte.

In höchsten Maße widersprüchlich erscheint jedoch die derzeitige Praxis (die im Rahmen von Fortbildungen von den Proponenten der 4.0 Skala vermittelt wird), wo die jeweils erzielten Leistungsfeststellungen der verschiedenen Kompetenzen zuerst mittels Mittelwertbildung als Durchschnitts-Score berechnet werden, um dann diese Scores gemäß der „möglichen Entscheidungsgrundlage“ nach Schlichtherle et al. (2013) in eine Ziffernote für das Jahreszeugnis zu transformieren. Da jedoch eine Mittelwertbildung einen Kompetenzverlauf nicht unbedingt sinnvoll abbilden kann, schlug bereits Marzano (2006), auf dessen Studien die 4.0 Skala basiert, eine Berechnung mittels „Powerlaw“ vor (siehe Beispiel auf S. 33 in diesem Positionspapier). In diesem

Zuge ergeht zweifache Kritik: Einerseits an die Proponenten der 4.0 Skala, die bei der Einführung der 4.0 Skala an Österreichischen NMS auf die Arbeiten von Marzano (2006) stark Bezug nehmen, ohne dann jedoch dessen sinnvolle Gedanken und Lösungsansätze hinsichtlich der formativen Leistungsbeurteilung mit Scores (Marzano 2006: Kapitel 6 ‚Final Scores and Grades‘ von S. 105-124) zu berücksichtigen. Vielmehr scheint es, dass von einer ungenannt gebliebenen anglo-amerikanischen Quelle (!) ein Beurteilungssystem für Österreichische NMS einfach übernommen wurde, ohne dessen Sinnhaftigkeit und Tauglichkeit zu überprüfen. Zweitens ergeht Kritik an ein Bildungs-System in Österreich, dass es zulässt, dass ein „unfertiges“ Produkt auf den Köpfen schulpflichtiger Kinder „ausprobiert“ bzw. "entwickelt" wird, ohne dass im Vorfeld eine wissenschaftliche Validierungsstudie stattgefunden hat.

„Kritisieren ist leicht und Kritisieren kann jeder“ bzw. „Wer kritisiert, soll es besser machen!“ Dieser Meinung könnte der eine oder andere Leser sein. Obwohl die Kritik eines Einzelnen gerne ungehört bleibt, weil sich einerseits Kritisierte „auf die Füße getreten fühlen“ und andererseits weil eine Systemänderung allenfalls von oben (und nicht von unten) erfolgen muss, kann evidenzbasierte Kritik, wie es der Anspruch im vorliegenden Positionspapier ist, jedoch auch neue Lösungswege eröffnen (siehe Kapitel 4).

Zusammenfassend ergeben sich nach Ansicht des Verfassers einige Eckpunkte anhand der in diesem Positionspapier dargestellten Analysen, wie eine Transformation von Scores in eine Ziffernote unter adäquater Berücksichtigung des Entwicklungsverlaufs der relevanten Kompetenzen pädagogisch sinnvoll erfolgen könnte bei gleichzeitiger Gewährleistung einer Beurteilungsgerechtigkeit.

Es bleibt zu hoffen, dass die Argumente in der vorliegenden kritischen Analyse der Leistungsbeurteilung mit der 4.0 Skala Gehör finden oder in gleicher Weise evidenzbasiert widerlegt werden.

Abschließend soll auf das Zitat einer Lehrperson verwiesen werden, mit der im Rahmen der Diskussionen hinsichtlich der Leistungsbeurteilung mit der 4.0 Skala gesprochen wurde, welches einerseits den Kern der Sache trifft und andererseits im wesentlichen Motor für die vorliegende Arbeit war:

**„Im Mittelpunkt stehen die Kinder!“**

## LITERATUR

- Eder F, Neuweg GH, Tonhauser J (2009) Leistungsfeststellung und Leistungsbeurteilung. In: Specht W (Hg.) Nationaler Bildungsbericht Österreich, Bd. 2. Analysen. Leykam, Graz: S. 247-268
- Marzano R (2006) Classroom Assessment and Grading that work. Association for Supervision and Curriculum Development, Alexandria, Virginia USA. pp. 182; online abrufbar: <http://ir.nmu.org.ua/bitstream/handle/123456789/139521/22ad1124ef267d1994032566a2beba72.pdf?sequence=1>
- Neuweg GH (2014) Schulische Leistungsbeurteilung. Rechtliche Grundlagen und pädagogische Hilfestellungen für die Schulpraxis. Trauner Verlag, Linz
- Neuweg GH (2009) Rechtsgrundlagen schulischer Leistungsbeurteilung. Problemzonen und Ansatzpunkte für Reformen. Schule & Recht; 1: 52-58
- Ruedl G (2015) In wie weit ist die derzeitig vorgeschlagene Entscheidungsgrundlage für die Umwandlung von 4.0 Scores in Ziffernnoten nach Schlichtherle et al. (2013) rechtskonform? Nicht veröffentl. Positionspapier
- Ruedl G (2015) Gibt es rechtskonforme Alternativen zur Mittelwertbildung bei der Ermittlung der Zeugnisnote? Nicht veröffentl. Positionspapier
- Schlichtherle B, Weiskopf-Prantner V, Westfall-Greiter (2013) Kriterienorientierte Leistungsfeststellung mit der 4.0 Skala
- Westfall-Greiter T (2014) Vorschläge für eine Entscheidungsgrundlage zur Ermittlung einer Gesamtnote auf Basis der Erprobung im SJ 2012/13
- Westfall-Greiter T (2013) Leistungsbeurteilung in der NMS: Geschärfte Konturen von Praxisproblemen. Erziehung und Unterricht, Nov-Dez 9-10: 1-9