



Universität Zürich
Institut für Bildungsevaluation

Vergleich der Resultate bei der Aufnahmeprüfung in die Gymnasien mit den Stellwerk-Testergebnissen

Kurzbericht zuhanden des Amts für Mittelschulen
des Kantons St. Gallen

Urs Moser
Zürich, 29. September 2009

Inhalt

1	Ausgangslage	3
2	Fragestellungen	4
3	Datengrundlage	4
4	Methode	6
5	Ergebnisse	8
5.1	Zusammenhänge zwischen den Ergebnissen verschiedener Beurteilungsinstrumente.....	8
	<i>Zusammenhang zwischen Prüfungsnoten und Stellwerk-Testergebnissen</i>	<i>8</i>
	<i>Zusammenhang zwischen Zeugnisnoten und Stellwerk-Testergebnissen</i>	<i>10</i>
	<i>Zusammenhang zwischen Empfehlung und Stellwerk-Testergebnissen.....</i>	<i>10</i>
	<i>Zusammenhang zwischen Zeugnisnoten und Prüfungsnoten</i>	<i>11</i>
	<i>Zusammenhang zwischen Empfehlung und Prüfungsnoten.....</i>	<i>11</i>
	<i>Überblick über die Zusammenhänge.....</i>	<i>11</i>
5.2	Empfehlungen von Lehrpersonen im Vergleich zu Zeugnisnoten, Prüfungsnoten und Testergebnissen	12
5.3	Erklärung der Prüfungsnote	15
5.4	Erklärung des Prüfungsentscheids	16
5.5	Stellwerk-Testergebnis oder Prüfungsnote?.....	17
6	Fazit.....	18
6.1	Zusammenfassung der wichtigsten Ergebnisse.....	18
6.2	Diskussion der Ergebnisse.....	18

1 Ausgangslage

Im Kanton St. Gallen ist für die Aufnahme ins Gymnasium eine Prüfung abzulegen. Zur Prüfung zugelassen sind Schülerinnen und Schüler, die die zweite Klasse der Oberstufe der Volksschule absolviert und höchstens das 17. Altersjahr erfüllt haben. Geprüft werden die Fächer Deutsch, Französisch und Mathematik. Die sprachlichen Prüfungen umfassen einen schriftlichen und einen mündlichen Teil, die Mathematikprüfung umfasst zwei schriftliche Prüfungsteile. Für die Zulassung zu den Schwerpunktfächern Latein, Musik oder Bildnerisches Gestalten sind zusätzliche Prüfungen beziehungsweise Eignungsabklärungen abzulegen¹.

Über die Aufnahme ins Gymnasium entscheidet die Prüfungspunktzahl, die sich aus der Summe der Noten in den Fächern Deutsch, Französisch sowie Mathematik Teil 1 und Mathematik Teil 2 ergibt. Wird zusätzlich die Lateinprüfung abgelegt, so führen je vier Fünftel der Noten zum Gesamtwert, der maximal 24 Punkte betragen kann. Wer 16 Punkte erreicht, wird ins Gymnasium aufgenommen, wobei auch Kandidatinnen und Kandidaten des zweiten Schuljahrs der Oberstufe mit 15 Punkten beziehungsweise des dritten Schuljahrs mit 15,5 Punkten aufgenommen werden können. Ausschlaggebend sind in diesen Fällen ein Eignungsbericht, die Dauer der Vorbildung und besondere Umstände².

Alle Schülerinnen und Schüler des Kantons St. Gallen, also auch die Kandidatinnen und Kandidaten der Gymnasien, bearbeiten in der zweiten Klasse der Oberstufe zugleich die Stellwerk-Tests. Stellwerk ist ein Testsystem des Lehrmittelverlags des Kantons St. Gallen, mit dem Kenntnisse in Deutsch, Französisch, Englisch, Mathematik sowie Natur und Technik in der 8. Klasse oder in der 9. Klasse der Volksschule am Computer überprüft werden. Stellwerk führt zu einem Stärken-Schwächen-Profil und kann auch für die Planung der Schul- und Berufswahl genutzt werden³.

Das Vorliegen von zwei Beurteilungen, die mit unabhängigen Instrumenten und in unterschiedlichen Kontexten zustande gekommen sind, führte in der kantonalen Rektorenkonferenz zur Frage, ob die Resultate in den Stellwerk-Tests einen Teil der Aufnahmeprüfung ersetzen könnten. Dadurch könnten Synergien genutzt und Ressourcen eingespart werden. Um diese Frage beantworten zu können, soll festgestellt werden, wieweit die Resultate der beiden Testmethoden korrelieren⁴.

¹ Aufnahmereglement des Gymnasiums sGS 215.31, Artikel 3 bis 9.

² Aufnahmereglement des Gymnasiums sGS 215.31, Artikel 18 bis 20.

³ www.stellwerk-check.ch

⁴ Auftrag für einen Vergleich der Resultate bei der Aufnahmeprüfung des Gymnasiums und dem Stellwerk8 der Volksschule, Schreiben des Amts für Mittelschulen vom 13. Februar 2009

2 Fragestellungen

Zur Beurteilung einer möglichen Funktion der Stellwerk-Testergebnisse im Rahmen des Aufnahmeverfahrens an die Gymnasien sollen folgende Fragen beantwortet werden⁵.

- Wie gut korrelieren die Ergebnisse der Aufnahmeprüfung in allen Prüfungsfächern mit denjenigen von Stellwerk im 8. Schuljahr?
- Wie gut stimmen die Zeugnisnoten der Oberstufe in verschiedenen Fächern mit den Ergebnissen von Stellwerk im 8. Schuljahr überein?
- Wie weit decken sich die Empfehlungen der Oberstufenlehrpersonen mit den Ergebnissen von Stellwerk im 8. Schuljahr und jenen der Aufnahmeprüfung?
- Welche durchschnittlichen Punktzahlen (Gesamtpunkte pro Fach) haben die in das Gymnasium aufgenommenen Kandidatinnen und Kandidaten bei Stellwerk im 8. Schuljahr erreicht? Diese Punktzahl könnte als Limite betrachtet werden.
- Wie gross ist der Anteil derjenigen, die aufgrund dieser Limite andere Entscheide betreffend die Aufnahme erzielt hätten (Aufnahme mit der Prüfung, jedoch nicht mit den Stellwerk-Testergebnissen oder umgekehrt)?

3 Datengrundlage

Für die Datenanalyse stehen Prüfungsergebnisse aus den Jahren 2007 und 2008 zur Verfügung. Tabelle 3.1 zeigt, von welcher Anzahl Schülerinnen und Schüler Informationen zur Aufnahmeprüfung vorliegen, aufgeteilt nach Schule und Prüfungserfolg.

Tabelle 3.1: Anzahl Schülerinnen und Schüler nach Schule und Prüfungserfolg

Schule	bestanden		nicht bestanden		Total
	n	%	n	%	n
Burggraben St. Gallen	257	75%	84	25%	341
Heerbrugg	157	84%	31	16%	188
Sargans	155	78%	43	22%	198
Wattwil	168	78%	47	22%	215
Wil	112	82%	24	18%	136
Total	849	79%	229	21%	1078

⁵ Auftrag für einen Vergleich der Resultate bei der Aufnahmeprüfung des Gymnasiums und dem Stellwerk8 der Volksschule, Schreiben des Amts für Mittelschulen vom 13. Februar 2009

Von 1078 Schülerinnen und Schülern liegen die Prüfungsergebnisse vor. 21 Prozent der Schülerinnen und Schüler bestanden die Prüfung nicht. Zusätzlich zu den Prüfungsergebnissen sind in der Regel auch folgende Angaben vorhanden:

- Zeugnisnoten der Oberstufe in Deutsch, Französisch und Mathematik
- Empfehlungen der Oberstufenlehrpersonen in Deutsch, Französisch und Mathematik
- Stellwerk-Testergebnisse in Deutsch, Französisch und Mathematik

Die *Zeugnisnoten* entsprechen einer wichtigen Information zur Beurteilung der Validität der Prüfungsergebnisse und der Stellwerk-Testergebnisse. Die Zeugnisnoten sollten für die Prüfungsergebnisse und die Stellwerk-Testergebnisse einen prognostischen Wert haben. Noten haben gegenüber Test- oder Prüfungsergebnissen den Vorteil, dass sie aufgrund einer Beobachtung und Beurteilung über einen längeren Zeitraum hinweg zustande kommen und dass auch fächerübergreifende Kompetenzen berücksichtigt werden können. Allerdings werden Noten in der Regel mit engem Bezug zum Klassenverband erteilt und lassen sich nicht über die Klassen hinweg vergleichen. Aufgrund der klassenbezogenen Beurteilungsmassstäbe sind deshalb keine allzu hohen Korrelationen zwischen Noten einerseits und Test- und Prüfungsergebnissen andererseits zu erwarten. Die Noten werden in Form von ganzen und halben Punkten zwischen 1 und 6 ausgewiesen.

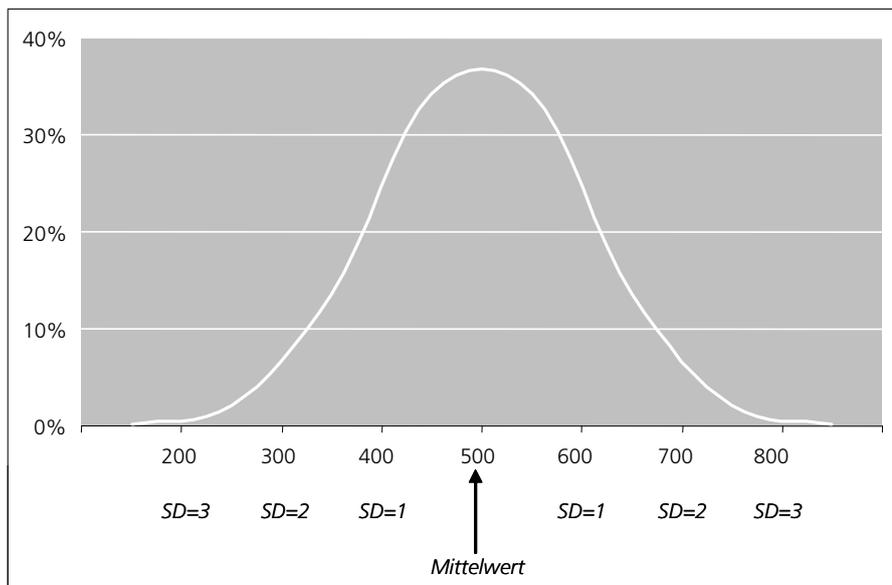
Die *Empfehlungen* der Oberstufenlehrpersonen werden in Grenzfällen zu einem Bestandteil des Aufnahmeverfahrens. Die Prüfung ist dann bestanden, wenn von 24 möglichen Punkten 16 erreicht werden, was einem Notendurchschnitt von 4,0 entspricht. Kandidatinnen und Kandidaten, die die Prüfung knapp nicht bestehen, können ebenfalls aufgenommen werden. Wer aus dem zweiten Schuljahr der Oberstufe an die Prüfung geht, braucht mindestens 15 Punkte, wer aus dem dritten Schuljahr aus der Oberstufe an die Prüfung geht, mindestens 15,5 Punkte. In diesen Fällen kann auf Empfehlungen der Lehrpersonen der Oberstufe zurückgegriffen werden. Die Empfehlungen werden in fünf Kategorien abgegeben: sehr gut, gut, bedingt, fraglich und nicht empfohlen. Von den Empfehlungen wird erwartet, dass sie besonders gut mit dem Erfolg der Schülerinnen und Schülern im Gymnasium übereinstimmen. Zudem wird auch ein enger Zusammenhang zwischen Empfehlung und Zeugnisnote erwartet. Beide Beurteilungen für ein Fach stammen von der gleichen Oberstufenlehrperson.

Die *Stellwerk-Testergebnisse* werden auf einer standardisierten Skala ausgewiesen, die von rund 200 Punkten bis 800 Punkte reicht. Der Mittelwert lag zum Zeitpunkt der ersten Durchführung, also bei der Normierung der Stellwerk-Skala durch die Schülerinnen und Schüler der Real- und Sekundarschulen des Kantons St. Gallen im Jahr 2005 (Referenzpopulation), bei $M = 500$ Punkten. Die Standardabweichung lag bei $SD = 100$ Punkten. Diese Kennzahlen können sich über die Jahre hinweg leicht ändern. Die Verteilung der Stellwerk-Testergebnisse entspricht einer Normalverteilung (Abbildung 3.1).

Die 2,3 Prozent der besten Schülerinnen und Schüler der Referenzpopulation erreichten Testergebnisse zwischen 700 und 800 Punkten, die 2,3 Prozent der schwächsten Schülerinnen und Schüler der Referenzpopulation erreichten Testergebnisse zwischen 200 und 300 Punkten. Rund 16 Prozent der besten Schülerinnen und Schüler erreichten

Testergebnisse, die über 600 Punkten liegen, rund 16 Prozent der schwächsten Schülerinnen und Schüler erreichten Testergebnisse, die unter 400 Punkten liegen. Rund 68 Prozent der Schülerinnen und Schüler erreichten Testergebnisse, die im mittleren Bereich zwischen 400 und 600 Punkten liegen.

Abbildung 3.1: Verteilung der Stellwerk-Testergebnisse



Die Stellwerk-Testergebnisse kommen gleich wie die Prüfungsergebnisse durch einen einmaligen Test zustande. Im Gegensatz zur Aufnahmeprüfung, die schriftlich zu lösen ist, wird der Stellwerk-Test ausschliesslich am Computer gelöst. Zudem kann es zwischen den Stellwerk-Tests und den Aufnahmeprüfungen auch inhaltliche Abweichungen geben. Es wird daher ein mittlerer Zusammenhang zwischen diesen beiden Datenquellen erwartet.

4 Methode

Für die Überprüfung der Zusammenhänge werden Korrelationskoeffizienten berechnet und zum Teil grafisch umgesetzt. Die Korrelation beschreibt die Beziehung von zwei Merkmalen, beispielsweise von den Prüfungsergebnissen und den Stellwerk-Testergebnissen in der Mathematik. Mit der Korrelationsrechnung wird ein Korrelationskoeffizient bestimmt, der angibt, ob die Annahme eines linearen Zusammenhangs zwischen zwei Merkmalen sinnvoll ist, und falls ja, wie stark diese Merkmale zusammenhängen. Eine plausible Annahme ist beispielsweise, dass die Prüfungsergebnisse in Mathematik mit den Stellwerk-Testergebnissen in Mathematik positiv korrelieren. Das heisst, dass beispielsweise ein hohes Prüfungsergebnis mit einem hohen Testergebnis einhergeht.

Der Zusammenhang zwischen den zwei Merkmalen wird durch den Korrelationskoeffizienten « r » beziffert. Der Koeffizient gibt die Stärke des Zusammenhangs zweier Merkmale in Form einer Zahl an. Er liegt zwischen -1 und $+1$. Ein positiver Wert bedeutet, dass ein hoher Wert eines Merkmals mit dem hohen Wert eines anderen Merkmals auftritt («je mehr, desto mehr»). Ein negativer Koeffizient besagt, dass ein hoher Wert mit dem niedrigen Wert eines anderen Merkmals auftritt («je mehr, desto weniger»). Ein Koeffizient von 0 bedeutet, dass zwischen den beiden Merkmalen kein Zusammenhang besteht. Korrelationskoeffizienten von $r = 0.1$ werden als klein beurteilt, solche von $r = 0.3$ werden als mittel beurteilt und solche von $r = 0.5$ werden als gross beurteilt.

Für die Beurteilung der Fähigkeiten in den Zeugnissen oder in der Aufnahmeprüfung wird die Notenskala genutzt. Die Stellwerk-Testergebnisse werden hingegen auf einer Skala von 200 bis 800 Punkten ausgewiesen. Die Notenskala führt zu Ergebnissen, die zwischen 1 und 6 liegen. Die Stellwerk-Skala führt zu Ergebnissen, die zwischen 200 und 800 Punkten liegen. Die beiden Skalen lassen sich nicht miteinander vergleichen. Bei beiden Skalen zeigt sich allerdings, dass sich die Verteilung der Ergebnisse stark einer Normalverteilung annähert. Die Ergebnisse lassen sich deshalb einfach in eine Standardnormalverteilung beziehungsweise in eine z -Skala transformieren.

Die z -Skala (Standardnormalverteilung) verfügt über spezifische Eigenschaften. So beträgt der Mittelwert $M = 0$ und die Standardabweichung $SD = 1$. Eine Einheit auf der z -Skala entspricht also einer Standardabweichung. Rund 68 Prozent der Schülerinnen und Schüler erreichen Ergebnisse, die zwischen -1 und $+1$ liegen, rund 95 Prozent erreichen Ergebnisse, die zwischen -2 und $+2$ liegen, und nahezu alle Ergebnisse liegen zwischen -3 und $+3$. Dank dieser Transformation können die Ergebnisse verschiedener Skalen miteinander verglichen werden.

5 Ergebnisse

5.1 Zusammenhänge zwischen den Ergebnissen verschiedener Beurteilungsinstrumente

Zusammenhang zwischen Prüfungsnoten und Stellwerk-Testergebnissen

Tabelle 5.1 enthält die Korrelationskoeffizienten für die Zusammenhänge zwischen den Noten in den Aufnahmeprüfungen und den Stellwerk-Testergebnissen. Der Koeffizient beträgt für Deutsch $r = .47$ Punkte, für Französisch $r = .52$ und für Mathematik $r = .58$. Die Zusammenhänge können als gross beurteilt werden. Je besser die Prüfungsergebnisse sind, desto besser sind auch die Stellwerk-Testergebnisse.

Tabelle 5.1: Zusammenhang Prüfungsnote – Stellwerk-Testergebnisse:
Korrelationskoeffizienten

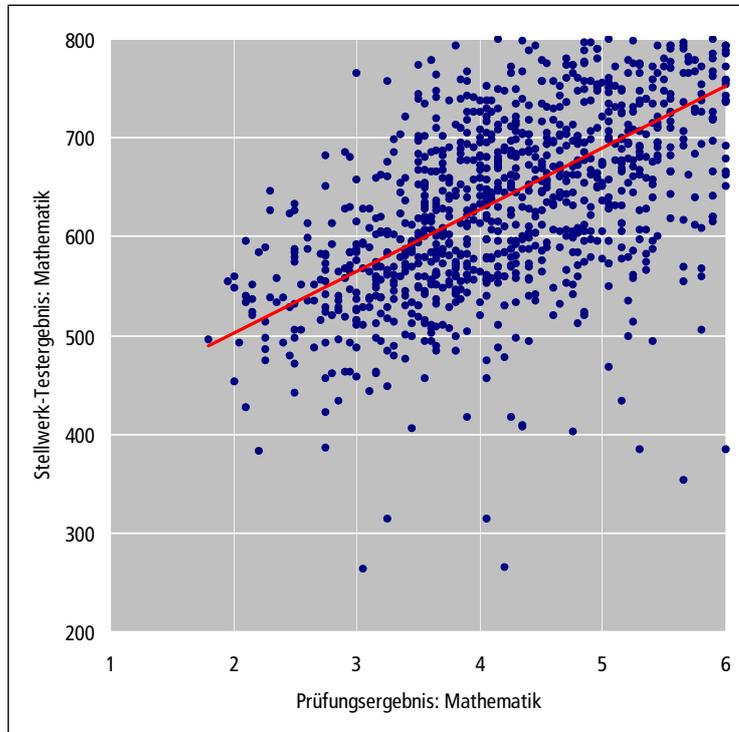
Prüfung	Stellwerk		
	Deutsch	Französisch	Mathematik
Deutsch	.47	.40	.19
Französisch	.36	.52	.18
Mathematik	.31	.21	.58
Mathematik Teil 1	.25	.18	.52
Mathematik Teil 2	.31	.19	.53

Abbildung 5.1 illustriert diesen grossen Zusammenhang zwischen den Noten in der Aufnahmeprüfung und den Stellwerk-Testergebnissen am Beispiel der Mathematik. Die Punkte in der Abbildung repräsentieren die Noten in der Aufnahmeprüfung und die Testergebnisse der Schülerinnen und Schüler. Die Position eines Punktes ergibt sich aus der Prüfungsnote (X-Achse) und der Punktzahl im Stellwerk-Test (Y-Achse).

Der lineare Zusammenhang zwischen den beiden Ergebnissen wird durch eine Gerade illustriert. Die Position der Datenpunkte zeigt, dass trotz engem Zusammenhang ein grosser Teil der Schülerinnen und Schüler je nach Beurteilungsinstrument verschieden abschneidet. Viele Datenpunkte liegen deshalb entweder deutlich über der Geraden (besseres Ergebnis im Stellwerk-Test als in der Aufnahmeprüfung) oder deutlich unter der Gerade (besseres Ergebnis in der Aufnahmeprüfung als im Stellwerk-Test).

Es gibt 229 Schülerinnen und Schüler, die im Stellwerk-Test zwischen 700 und 800 Punkte erreichten. Von diesen 229 Schülerinnen und Schülern erreichten rund 15 Prozent in der Aufnahmeprüfung nur eine ungenügende Note, rund 40 Prozent eine Note zwischen 4 und 5 und rund 45 Prozent eine Note von 5 oder mehr.

Abbildung 5.1: Zusammenhang Prüfungsnoten – Stellwerk-Testergebnisse: Datenpunkte der Schülerinnen und Schüler am Beispiel der Mathematik



Demgegenüber erreichten 421 Schülerinnen und Schüler eine ungenügende Note in der Aufnahmeprüfung Mathematik. Von diesen erreichten rund 8 Prozent ein Stellwerk-Testergebnis, das über 700 Punkten liegt, 63 Prozent ein Stellwerk-Testergebnis, das unter 600 Punkten liegt, und rund 29 Prozent ein Stellwerk-Testergebnis, das zwischen 600 und 700 Punkten liegt.

Wie sind diese Zusammenhänge zu interpretieren und was würde es für die Schülerinnen und Schüler bedeuten, wenn anstelle der Aufnahmeprüfung die Stellwerk-Testergebnisse genutzt würden?

Die Punkte in Abbildung 5.1 zeigen, dass sich die Ergebnisse aufgrund unterschiedlicher Beurteilungsinstrumente (Test oder Prüfung) zum Teil deutlich unterscheiden. Ein Korrelationskoeffizient von $r = .58$ bedeutet, dass rund ein Drittel der Leistungsunterschiede in der Prüfungsnote durch das Stellwerk-Testergebnis erklärt werden kann. Die übrigen zwei Drittel der Leistungsunterschiede müssen durch andere Merkmale erklärt werden. Der Zusammenhang zwischen der Prüfungsnote und dem Stellwerk-Testergebnis von $r = 0.58$ ist zwar relativ gross, jedoch weit weg von einem perfekten linearen Zusammenhang und ermöglicht keine präzise Prognose. Die Note in der Aufnahmeprüfung kann durch das Stellwerk-Testergebnis für zwei Drittel der Schülerinnen und Schüler nicht genau vorhergesagt werden.

Zusammenhang zwischen Zeugnisnoten und Stellwerk-Testergebnissen

Tabelle 5.2 enthält die Korrelationskoeffizienten der Zusammenhänge zwischen den Zeugnisnoten und den Stellwerk-Testergebnissen. Die Zusammenhänge können als mittel bis gross beurteilt werden. Der Korrelationskoeffizient beträgt für Deutsch $r = .44$, für Französisch $r = .37$ und für Mathematik $r = .47$. Die Koeffizienten sind im Vergleich zu jenen der Zusammenhängen zwischen Prüfungsergebnissen und Stellwerk-Testergebnissen leicht geringer. Wiederum lässt sich der engste Zusammenhang für Mathematik, der geringste für Französisch nachweisen.

Tabelle 5.2: Zusammenhang Zeugnisnoten – Stellwerk-Testergebnisse:
Korrelationskoeffizienten

Zeugnis	Stellwerk		
	Deutsch	Französisch	Mathematik
Deutsch	.44	.36	.27
Französisch	.25	.37	.21
Mathematik	.28	.21	.47

Zusammenhang zwischen Empfehlung und Stellwerk-Testergebnissen

Tabelle 5.3 enthält die Korrelationskoeffizienten der Zusammenhänge zwischen den Empfehlungen fürs Gymnasium durch die Oberstufenlehrpersonen und den Stellwerk-Testergebnissen. Die Zusammenhänge sind etwas geringer als jene zwischen den Prüfungsnoten und den Stellwerk-Testergebnissen und ähnlich gross wie jene zwischen den Zeugnisnoten und den Stellwerk-Testergebnissen. Der Korrelationskoeffizient beträgt für Deutsch $r = .39$, für Französisch $r = .36$ und für Mathematik $r = .48$.

Tabelle 5.3: Zusammenhang Empfehlung – Stellwerk-Testergebnisse:
Korrelationskoeffizienten

Empfehlung	Stellwerk		
	Deutsch	Französisch	Mathematik
Deutsch	.39	.37	.33
Französisch	.34	.36	.35
Mathematik	.35	.24	.48

Zusammenhang zwischen Zeugnisnoten und Prüfungsnoten

Tabelle 5.4 enthält die Korrelationskoeffizienten für die Zusammenhänge zwischen den Zeugnisnoten und den Prüfungsnoten. Der Zusammenhang beträgt für Deutsch $r = .54$ Punkte, für Französisch $r = .51$ und für Mathematik $r = .51$. Prüfungsnoten und Zeugnisnoten hängen insgesamt etwa gleich stark zusammen wie Prüfungsnoten und Stellwerk-Testergebnisse.

Tabelle 5.4: Zusammenhang Zeugnisnoten – Prüfungsnoten: Korrelationskoeffizienten

Zeugnis	Prüfung		
	Deutsch	Französisch	Mathematik
Deutsch	.54	.44	.26
Französisch	.38	.51	.15
Mathematik	.19	.18	.51

Zusammenhang zwischen Empfehlung und Prüfungsnoten

Zur Beurteilung des Zusammenhangs zwischen den Prüfungsnoten und den Stellwerk-Testergebnissen wurde auch der Zusammenhang zwischen den Empfehlungen und den Prüfungsnoten berechnet. Die Koeffizienten sind in Tabelle 5.5 enthalten. Der Zusammenhang beträgt für Deutsch $r = .49$ Punkte, für Französisch $r = .49$ und für Mathematik $r = .55$. Die Empfehlungen korrelieren etwa ähnlich stark mit den Prüfungsnoten wie mit den Zeugnisnoten.

Tabelle 5.5: Zusammenhang Empfehlung – Prüfungsnoten: Korrelationskoeffizienten

Empfehlung	Prüfung		
	Deutsch	Französisch	Mathematik
Deutsch	.49	.51	.36
Französisch	.43	.49	.34
Mathematik	.30	.29	.55

Überblick über die Zusammenhänge

Tabelle 5.6 gibt einen Überblick über die Zusammenhänge zwischen den Beurteilungen, die aufgrund unterschiedlicher Instrumente zustande gekommen sind. Insgesamt sind die Unterschiede zwischen den Korrelationskoeffizienten eher gering.

Tabelle 5.6 Überblick über die Zusammenhänge nach Instrument und Fach

	Stellwerk-Testergebnisse		
	Deutsch	Französisch	Mathematik
Prüfungsnote	.47	.52	.58
Zeugnisnote	.44	.37	.47
Empfehlung	.39	.36	.48
	Prüfungsnote		
	Deutsch	Französisch	Mathematik
Zeugnisnote	.54	.51	.51
Empfehlung	.49	.49	.55

Der geringste Zusammenhang zeigt sich zwischen dem Stellwerk-Testergebnis und der Empfehlung für Französisch ($r = .36$), der engste Zusammenhang zwischen dem Stellwerk-Testergebnis und der Prüfungsnote für Mathematik ($r = .58$). Die Zusammenhänge zwischen den einzelnen Beurteilungsquellen und den Stellwerk-Testergebnissen sind für die Mathematik etwas höher als für die Sprachen. Die Stellwerk-Testergebnisse korrelieren am stärksten mit den Prüfungsnoten. Der Grund könnte darin liegen, dass in der Empfehlung und in der Zeugnisnote andere Informationen in die Beurteilung einfließen als mit Tests gemessen werden. Empfehlung und Zeugnisnoten basieren auf den Beurteilungen über einen längeren Zeitraum hinweg und beziehen sich auch auf produktive Kompetenzen wie Sprechen und Texteschreiben, die mit der Prüfung und dem Stellwerk-Test nicht erfasst werden. Dies wirkt sich für die Beurteilung der Sprachen etwas stärker aus als für die Beurteilung der Mathematik.

Die Prüfungsnoten hängen etwas stärker mit den Zeugnisnoten und den Empfehlungen zusammen als die Stellwerk-Testergebnisse. Eher gering ist dieser Unterschied in der Mathematik, etwas grösser sind die Unterschiede in Deutsch und Französisch.

5.2 Empfehlungen von Lehrpersonen im Vergleich zu Zeugnisnoten, Prüfungsnoten und Testergebnissen

Die Zeugnisnoten und Empfehlungen werden von den Lehrpersonen der Oberstufe vergeben. Im Vergleich zu den Prüfungsergebnissen oder zu den Ergebnissen im Stellwerk-Test stützen sich Zeugnisnoten und Empfehlungen auf weit mehr Kenntnisse und Fähigkeiten, als mit einer einmaligen Prüfung oder mit einem einmaligen Test erfasst werden können. Noten und Empfehlungen entstehen über einen bestimmten Zeitraum hinweg. Sie beziehen sich zudem eher auf die Leistungsverteilung der Mitschülerinnen und Mitschüler und entsprechen somit einer Rangreihe innerhalb der Klasse. Der Bezug zu expliziten Kompetenzbeschreibungen fehlt in der Regel.

Abbildung 5.2 zeigt die durchschnittlichen Ergebnisse der Schülerinnen und Schüler für jede Empfehlungskategorie bei den Zeugnisnoten, in der Aufnahmeprüfung sowie im Stellwerk-Test für die Mathematik. Damit die Ergebnisse verglichen werden können, wurden die drei Quellen in die z-Skala mit Mittelwert $M = 0$ und Standardabweichung $SD = 1$ transformiert (vergleiche Kapitel 4, Methode).

Abbildung 5.2: Empfehlung, Zeugnisnote, Prüfungsnote und Stellwerk-Testergebnis: Mathematik

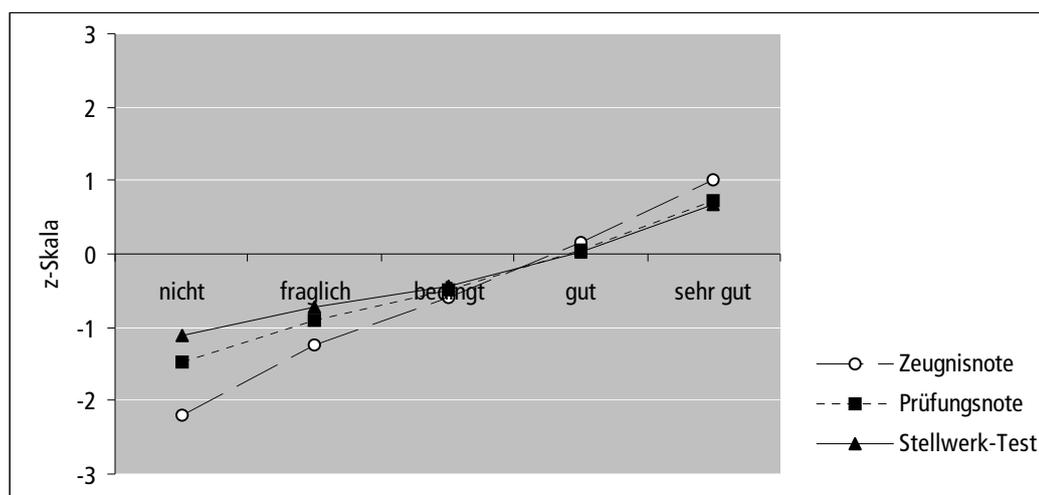


Abbildung 5.2 zeigt, dass die Empfehlungen nicht nur mit den Zeugnisnoten, sondern auch mit den Prüfungsnoten und den Stellwerk-Testergebnissen einhergehen. Die Abstufung der Empfehlungen in der Mathematik wird durch die Zeugnisnoten zwar am besten wiedergegeben. Die Prüfungsergebnisse und die Stellwerk-Testergebnisse vermögen aber ebenfalls zwischen den fünf Kategorien entsprechend den Erwartungen zu differenzieren. Die Prüfungsnote spiegelt die Empfehlungen bei den Kategorien «nicht» und «fraglich» etwas besser als das Stellwerk-Testergebnis.

Abbildung 5.3 zeigt die durchschnittlichen Ergebnisse der Schülerinnen und Schüler für jede Empfehlungskategorie bei den Zeugnisnoten, in der Aufnahmeprüfung sowie im Stellwerk-Test für Deutsch. Auch für Deutsch wird die Abstufung der Empfehlungen am besten von der Zeugnisnote abgebildet. Im Gegensatz zu den Ergebnissen für die Mathematik differenziert allerdings die Prüfungsnote zwischen den beiden Kategorien «nicht» und «fraglich» besser als das Stellwerk-Testergebnis. Die Aufnahmeprüfung bildet die erwarteten Leistungen bei den schwächeren Schülerinnen und Schüler besser ab als das Stellwerk-Testergebnis. Bei den besseren Schülerinnen und Schülern führen die Aufnahmeprüfungen und die Stellwerk-Testergebnisse hingegen zu einer nahezu vergleichbaren Differenzierung wie die Zeugnisnote.

Abbildung 5.3: Empfehlung, Zeugnisnote, Prüfungsnote und Stellwerk-Testergebnis:
Deutsch

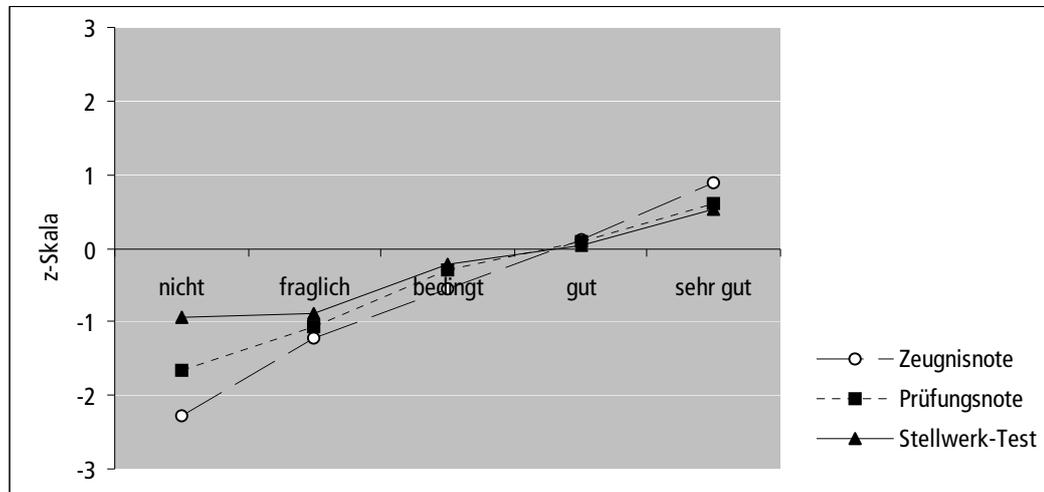
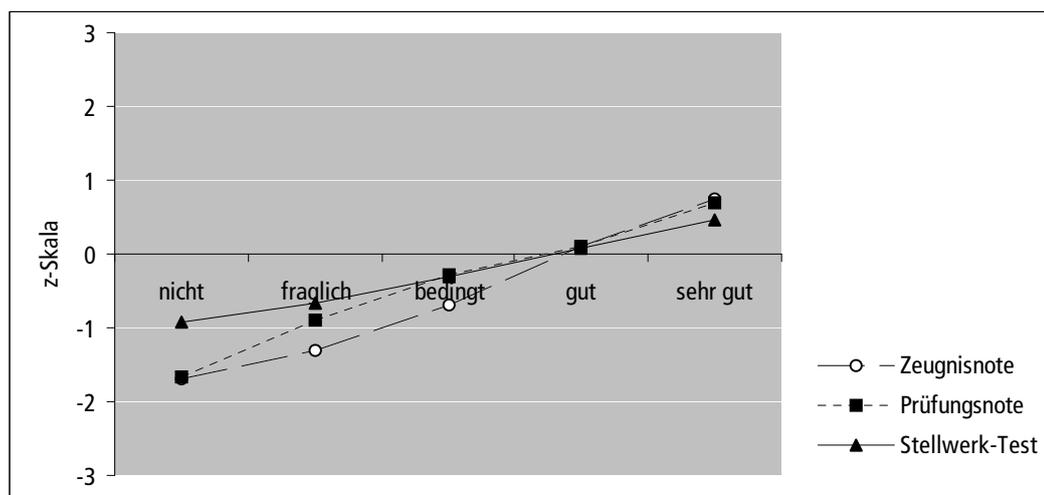


Abbildung 5.4 zeigt die durchschnittlichen Ergebnisse der Schülerinnen und Schüler für jede Empfehlungskategorie bei den Zeugnisnoten, in der Aufnahmeprüfung sowie im Stellwerk-Test für Französisch. Die Abstufung der Empfehlungen in Französisch wird wiederum von der Zeugnisnote am besten abgebildet, während mit den Stellwerk-Testergebnissen etwas weniger klar vorhergesagt werden kann, wer zu den verschiedenen Kategorien gehört.

Abbildung 5.4: Empfehlung, Zeugnisnote, Prüfungsnote und Stellwerk-Testergebnis:
Französisch



5.3 Erklärung der Prüfungsnote

Die Darstellung der Korrelationen hat gezeigt, dass es zum Teil enge Zusammenhänge zwischen Prüfungsnoten, Zeugnisnoten und Stellwerk-Testergebnissen gibt. Die meisten Korrelationen sind mittel oder gross und liegen zwischen $r = .35$ und $r = .55$. Durch eine Regressionsanalyse wird nun gezeigt, welches der drei Merkmale (Zeugnisnote, Empfehlung, Stellwerk-Testergebnis) den Erfolg bei der Prüfung am besten zu erklären vermag. Statt die Korrelation zwischen zwei Merkmalen zu berechnen, wird bei einer Regressionsanalyse eine eindeutige Richtung des Zusammenhangs angenommen. Mit der Regressionsanalyse wird die Abhängigkeit des Prüfungserfolgs von Zeugnisnoten, Empfehlungen und Stellwerk-Testergebnissen untersucht. Dadurch wird aufgezeigt, wie gut sich der Prüfungserfolg durch die Zeugnisnote, die Empfehlung der Oberstufenlehrperson und das Stellwerk-Testergebnis erklären lässt.

Tabelle 5.7 enthält die Ergebnisse der Regressionsanalysen, die für die drei Fachbereiche Deutsch, Französisch und Mathematik getrennt durchgeführt wurden. Damit die Bedeutung der drei Bestandteile für die Prüfungsnote verglichen werden kann, wurden wiederum die standardisierten z-Werte genutzt. Die Werte in der Spalte «Beta» sind wie z-Werte zu interpretieren. Sie zeigen, um wie viele z-Punkte die Prüfungsnote ansteigt, wenn die Zeugnisnote, die Empfehlung oder das Stellwerk-Testergebnis um eine z-Einheit (eine Standardabweichung) ansteigt. Der p-Wert zeigt, bei welcher Irrtumswahrscheinlichkeit das Ergebnis statistisch signifikant ist.

Tabelle 5.7: Erklärung der Prüfungsnote durch Zeugnisnote, Empfehlung und Stellwerk-Testergebnisse

	Deutsch		Französisch		Mathematik	
	Beta	p-Wert	Beta	p-Wert	Beta	p-Wert
Zeugnisnote	0.31	0.00	0.18	0.00	0.16	0.00
Empfehlung	0.17	0.00	0.23	0.00	0.24	0.02
Stellwerk-Testergebnis	0.26	0.00	0.38	0.00	0.39	0.00

Bei einem Anstieg der Zeugnisnote für Deutsch um eine Standardabweichung steigt die Prüfungsnote um 0.31 Standardabweichungen. Bei einem Anstieg der Empfehlung für Deutsch um eine Standardabweichung steigt die Prüfungsnote um 0.17 Standardabweichungen. Bei einem Anstieg des Stellwerk-Testergebnisses für Deutsch um eine Standardabweichung steigt die Prüfungsnote um 0.26 Standardabweichungen. Für Deutsch wird die Prüfungsnote am besten durch die Zeugnisnote vorausgesagt. Für Französisch und Mathematik wird die Prüfungsnote hingegen am besten durch die Stellwerk-Testergebnisse vorausgesagt. Die Bedeutung der Zeugnisnote und der Empfehlung ist für Französisch und Mathematik deutlich geringer.

5.4 Erklärung des Prüfungsentscheids

Der hohe Vorhersagewert der Stellwerk-Testergebnisse für die *Prüfungsnote* heisst nicht zwingend, dass die Stellwerk-Testergebnisse auch den *Prüfungsentscheid* – bestanden oder nicht bestanden – am besten vorhersagen. Tabelle 5.8 zeigt die Differenzen in den Stellwerk-Testergebnissen der Kandidatinnen und Kandidaten nach Prüfungserfolg (bestanden – nicht bestanden). Die Differenzen der Mittelwerte der Stellwerk-Testergebnisse sind für Französisch und Deutsch ähnlich gross, für Mathematik deutlich grösser.

Tabelle 5.8: Stellwerk-Testergebnisse nach Prüfungsentscheid

Stellwerk-Testergebnis	bestanden	nicht bestanden	Differenzen
Deutsch	597	545	52
Französisch	599	546	53
Mathematik	668	572	96

Auch für die Beurteilung des Zusammenhangs zwischen Prüfungsentscheid und Stellwerk-Testergebnissen lohnt es sich, den Vergleich zwischen den beiden Gruppen «bestanden» und «nicht bestanden» anhand der standardisierten Ergebnisse (z-Skala) durchzuführen. Im Gegensatz zu den Angaben in Tabelle 5.8 sind in Tabelle 5.9 die durchschnittlichen Ergebnisse der Schülerinnen und Schüler, die die Prüfung bestanden oder nicht bestanden haben, als standardisierte z-Werte angegeben. Dadurch lassen sich die Differenzen zwischen diesen beiden Gruppen direkt miteinander vergleichen.

Tabelle 5.9: Prüfungsentscheid nach Beurteilungsquelle

Stellwerk-Testergebnis	bestanden	nicht bestanden	Differenzen
<i>Deutsch</i>			
Zeugnisnote	0.17	-0.60	0.77
Empfehlung	0.22	-0.79	1.01
Stellwerk-Testergebnisse	0.17	-0.63	0.80
<i>Französisch</i>			
Zeugnisnote	0.14	-0.53	0.67
Empfehlung	0.20	-0.72	0.92
Stellwerk-Testergebnisse	0.15	-0.55	0.71
<i>Mathematik</i>			
Zeugnisnote	0.18	-0.65	0.83
Empfehlung	0.23	-0.79	1.03
Stellwerk-Testergebnisse	0.20	-0.72	0.91

Die grösste Differenz zwischen den beiden Gruppen (bestanden – nicht bestanden) besteht jeweils in der Empfehlung durch die Lehrpersonen. Sie beträgt für Deutsch 1.01 Punkte, für Französisch 0.92 Punkte und für Mathematik 1.03 Punkte auf der z-Skala. Das heisst, dass die Lehrpersonen am besten zwischen erfolgreichen und erfolglosen Kandidatinnen und Kandidaten unterscheiden können. Auch die Zeugnisnoten und die Stellwerk-Testergebnisse trennen deutlich zwischen diesen beiden Gruppen, jedoch weniger stark. Die Differenz zwischen erfolglosen Kandidatinnen und erfolgreichen Kandidaten ist für alle Fächer in den Stellwerk-Testergebnissen leicht grösser als die Differenz in den Zeugnisnoten.

5.5 Stellwerk-Testergebnis oder Prüfungsnote?

Um zu zeigen, für wie viele Schülerinnen und Schüler die Verwendung der Stellwerk-Testergebnissen anstelle der Prüfungsnoten zu einem anderen Prüfungsentscheid geführt hätte, wurde die Erfolgsquote von 21,2 Prozent dazu benutzt, jenen durchschnittlichen Stellwerk-Testwert zu bestimmen, bei dem ebenfalls 21,2 Prozent der Kandidatinnen und Kandidaten die Aufnahmeprüfung nicht bestanden hätten und 78,8 Prozent die Aufnahmeprüfung bestanden hätten (Punktzahl als Limite für den Aufnahmeentscheid). Der durchschnittliche Stellwerk-Testwert entspricht dem Mittelwert der Testergebnisse in Deutsch, Französisch und Mathematik. Er beträgt 555 Punkte. Das heisst, dass 21,2 Prozent der Kandidatinnen und Kandidaten einen durchschnittlichen Stellwerk-Testwert erreicht hatten, der bei 555 Punkten oder darunter lag. 78,8 Prozent der Kandidatinnen und Kandidaten erreichten einen durchschnittlichen Stellwerk-Testwert, der über 555 Punkten lag. Tabelle 5.10 zeigt, für wie viele Kandidatinnen und Kandidaten die Nutzung des durchschnittlichen Stellwerk-Testergebnisses zu einem anderen Prüfungsentscheid geführt hätte.

Tabelle 5.10: Änderung des Prüfungsergebnisses durch Nutzung der Stellwerk-Tests

	Anzahl	Anteil
zweimal bestanden	737	68.6%
zweimal nicht bestanden	117	10.9%
einmal bestanden, einmal nicht bestanden	221	20.6%
Total	1075	100%

Für rund vier Fünftel der Schülerinnen und Schüler würde sich der Prüfungsentscheid nicht ändern. Sie würden ohnehin die Prüfung bestehen oder nicht bestehen. Für rund ein Fünftel der Schülerinnen und Schüler hingegen würde der Entscheid anders ausfallen. Die eine Hälfte würde von den Stellwerk-Tests profitieren und ins Gymnasium aufgenommen, für die andere Hälfte wäre es ein Nachteil, wenn die Stellwerk-Testergebnisse genutzt würden, weil sie nicht ins Gymnasium aufgenommen würden.

6 Fazit

6.1 Zusammenfassung der wichtigsten Ergebnisse

Im Kanton St. Gallen ist der Übertritt in die Gymnasien durch eine schriftliche Prüfung in den Fächern Deutsch, Französisch und Mathematik geregelt. Der grosse Teil der Schülerinnen und Schüler, die eine Prüfung fürs Gymnasium ablegen, hat im letzten Schuljahr der Oberstufe für die Beurteilung der Fähigkeiten in Deutsch, Französisch und Mathematik auch die Stellwerk-Tests bearbeitet. Mit einer statistischen Analyse wurde deshalb überprüft, inwieweit die Prüfungsergebnisse für die Gymnasien mit den Ergebnissen der Stellwerk-Tests übereinstimmen.

Die Analysen zeigen, dass die Prüfungsnoten mit den Stellwerk-Testergebnissen zum Teil stark zusammenhängen. Eine präzise Vorhersage der Prüfungsnote lässt sich durch die Stellwerk-Testergebnisse jedoch nicht machen. Allerdings hängen auch die Zeugnisnoten oder die Empfehlungen der Lehrpersonen nicht etwa stärker mit der Prüfungsnote zusammen. Im Gegenteil: Ein Vergleich der Zusammenhänge zwischen den Ergebnissen verschiedener Beurteilungsinstrumente (Prüfungsnote, Zeugnisnote, Empfehlung fürs Gymnasium, Stellwerk-Testergebnis) zeigt, dass die Prüfungsnote in den Fächern Französisch und Mathematik durch die Stellwerk-Testergebnisse am besten vorausgesagt wird. Die Prüfungsnote im Fach Deutsch wird durch das Stellwerk-Testergebnis ebenfalls gut vorausgesagt, jedoch noch besser durch die Zeugnisnote und die Empfehlung der Oberstufenlehrperson.

Die Analysen zeigen allerdings auch, dass der Prüfungsentscheid (bestanden oder nicht bestanden) am besten durch die Empfehlungen der Lehrpersonen vorausgesagt wird. Allerdings hängen auch die Stellwerk-Testergebnisse relativ eng mit dem Prüfungsentscheid zusammen. Die Abweichungen zwischen den Beurteilungen durch die Lehrpersonen (Empfehlungen und Zeugnisnoten) und den Stellwerk-Testergebnissen sind generell nicht sonderlich gross. Einzig im Fach Deutsch gelingt für Kandidatinnen und Kandidaten, die fürs Gymnasium nicht oder nur bedingt empfohlen sind, die erwartete Einteilung mittels Stellwerk-Testergebnissen nicht. Die Stellwerk-Testergebnisse dieser Kandidatinnen und Kandidaten sind besser, als sie aufgrund der Zeugnis- oder Prüfungsnoten erwartet werden.

Wären anstelle der Prüfungsnoten die Stellwerk-Testergebnisse genutzt worden, dann wäre der Prüfungsentscheid für vier Fünftel der Kandidatinnen und Kandidaten gleich ausgefallen. Beim restlichen Fünftel der Schülerinnen und Schüler wäre der Prüfungsentscheid hingegen positiv statt negativ oder negativ statt positiv ausgefallen.

6.2 Diskussion der Ergebnisse

Die Frage, ob die Aufnahmeprüfung durch die Stellwerk-Tests ersetzt werden kann, lässt sich aufgrund der vorliegenden Ergebnisse durchaus positiv beantworten. Für vier Fünftel der Schülerinnen und Schüler würde sich vermutlich nichts ändern. Für die nicht oder nur bedingt empfohlenen Schülerinnen und Schüler wäre der Stellwerk-Test eher ein Vorteil, für Schülerinnen und Schüler, deren Empfehlungen eher positiv ausfallen,

die aber in den Stellwerk-Tests eher schlecht abschneiden, wäre der Stellwerk-Test gegenüber der bisherigen Prüfung eher ein Nachteil.

Viel wichtiger für die Beantwortung der Frage, ob die Aufnahmeprüfung durch die Stellwerk-Tests ersetzt werden kann, ist unseres Erachtens die Bedeutung der Stellwerk-Tests im Rahmen der Qualitätssicherung der Schulen des Kantons St. Gallen. Inwiefern es sinnvoll ist, die Stellwerk-Tests für Selektionsentscheide zu nutzen, hängt insbesondere von den Zielen dieser Tests ab. Vom Lehrmittelverlag des Kantons St. Gallen wird das Ziel der Stellwerk-Tests wie folgt umschrieben:

«Stellwerk ermöglicht jedem Schüler und jeder Schülerin eine individuelle Analyse des Leistungsstandes in den fünf Fachbereichen Mathematik, Deutsch, Natur und Technik, Englisch und Französisch. Das Leistungsprofil weist die Stärken und Schwächen des Lernenden aus und hilft die richtigen Massnahmen für eine wirkungsvolle Förderung zu treffen. Mit dem Referenzrahmen und den darin aufgeführten Kann-Formulierungen können sich alle Beteiligten Transparenz über die Anforderungen verschaffen. Das Leistungsprofil und der Analyseraster unterstützen die Beteiligten, geeignete Massnahmen zur Förderung der einzelnen Schüler und Schülerinnen zu treffen.»⁶

Bis anhin ist das Ziel der Stellwerk-Tests folglich, den Lernenden ein standardisiertes Leistungsprofil zu ermöglichen, das ihnen hilft, die richtigen Massnahmen für eine Förderung zu treffen. Die Tests dienen primär der individuellen Förderung. Werden die Stellwerk-Tests in Zukunft auch für Selektionsentscheide eingesetzt, dann sollen mit einem Instrument zwei Funktionen erfüllt werden: Fördern und Selektionieren. Genau darin liegt ein Problem. Verschiedene Studien zeigen, dass über Tests die individuelle Förderung durchaus positiv beeinflusst werden kann. Tests können zu notwendigen Informationen für die individuelle Ausrichtung des Lernens im Unterricht führen. Tests können zudem auch erfolgreich für Selektionsentscheide und Schulabschlüsse genutzt werden. Dass der gleiche Test sowohl für die individuelle Förderung als auch für Selektionsentscheide sinnvoll genutzt werden kann, wird hingegen bezweifelt.

Testergebnisse sind für Selektionsentscheide nur dann geeignete Masse, wenn die Schülerinnen und Schüler auf die getesteten Inhalte vorbereitet werden. Dies wird im Falle der Nutzung der Stellwerk-Tests für Selektionsentscheide als Nachteil betrachtet. Zum einen kann die Nutzung von Testsystemen für Selektionsentscheide tief greifende Auswirkungen auf den Unterricht haben und das Curriculum beziehungsweise den Unterricht einengen. Dies wird im Fall der Stellwerk-Tests vor allem deshalb als problematisch eingestuft, weil die Tests ausschliesslich produktive Kenntnisse am Computer abfragen. Zum andern zeigen Erfahrungen, dass die Verbindung von Testergebnissen mit Selektionsentscheiden dazu führt, dass Qualitätssysteme unterlaufen werden. Konkret ist zu befürchten, dass durch die Erhöhung der Bedeutung der Stellwerk-Tests für die Schullaufbahn die Testaufgaben nicht geheim gehalten werden könnten. Erfahrungen zeigen, dass sich Schülerinnen und Schüler dann nicht auf die getesteten Inhalte

⁶ www.stellwerk-check.ch

und Kompetenzen, sondern auf Testaufgaben und spezifische Aufgabenformate vorbereiten. Zudem würde die Vorbereitung der Schülerinnen und Schüler auf die Stellwerk-Tests und somit die Fokussierung auf reproduzierbares Wissen zur Pflicht. Ob dies im Sinne einer gymnasialen Bildung ist, sollte zumindest diskutiert werden.

Selbstverständlich hat auch die reguläre Aufnahmeprüfung gleiche Folgen. Was geprüft wird, muss vermittelt werden und bestimmt meist nicht nur die Prüfungsvorbereitung, sondern hat generell Auswirkungen auf den Unterricht. Aus dieser Perspektive beurteilt gibt es zwischen Stellwerk und der Aufnahmeprüfung keine grossen Unterschiede. Die Aufnahmeprüfung wird allerdings jedes Jahr neu entwickelt und vor allem liegt beispielsweise im Fach Deutsch das Schwergewicht unter anderem auch auf dem sprachlichen Ausdruck beim Verfassen von Texten, beim Verständnis und der Interpretation von stufengemässen Texten sowie bei der schriftlichen Darstellung von Gelesenem und Beobachtbarem; alles Bereiche, die mit einem computerbasierten Testsystem nicht erfasst werden können.

Gegen den Einsatz von Stellwerk im Rahmen von Aufnahmeprüfungen sprechen vermutlich auch juristische Überlegungen. Zum einen sind die Stellwerk-Tests für die Kandidatinnen und Kandidaten unterschiedlich. Die Schülerinnen und Schüler bearbeiten eine spezifische Anzahl von Aufgaben, die auf ihre Fähigkeiten am besten zugeschnitten sind. Dies hat zur Folge, dass jede Schülerin und jeder Schüler einen anderen Test bearbeitet. Die Prüfung besteht zwar auf einem geeichten Aufgabenpool, jedoch aus unterschiedlichen Tests. Zum andern müssen die Aufgaben in einem adaptiven Testsystem geheim gehalten werden, weil sie immer wieder eingesetzt werden. Ein Einsichtsrecht könnte den Kandidatinnen und Kandidaten sowie deren Eltern deshalb kaum im gewünschten Sinne gewährt werden.

Im Vergleich zu den Stellwerk-Testergebnissen ist die Funktion bei einer Aufnahmeprüfung eindeutig und unmissverständlich bestimmt. Ziel ist eine Standardisierung von Selektionsentscheiden. Sofern die Stellwerk-Tests in Zukunft die gleiche Funktion einnehmen sollen, ist es gegenüber den Schülerinnen und Schülern sowie ihren Eltern, aber auch gegenüber den Kunden des Lehrmittelverlags St. Gallen redlich, dies zumindest frühzeitig offen zu kommunizieren.