

Das Schätzen von Größen – Welche Antwort ist angemessen?

Beim Schätzen hängt es maßgeblich vom Ziel ab, welche Genauigkeit der zu ermittelnde Wert aufweisen muss. Während im Alltag der situative Kontext “beurteilt”, inwieweit es sich um ein adäquates Schätzergebnis handelt, fehlt ein derartiges Kriterium im Unterricht (Forrester et al., 1990; Ruwisch et al., 2015). Wann kann oder sollte im Unterricht oder in schriftlichen Tests somit von einer guten, wann von einer weniger guten oder gar unpassenden Schätzung gesprochen werden? Zwar liegen sowohl in der mathematikdidaktischen als auch psychologischen Forschung quantitative Daten zum Schätzen von Anzahlen und Größen sowie zum überschlagenden Rechnen vor, die Beurteilung der Schätzgenauigkeit wird jedoch höchst unterschiedlich gehandhabt.

Angemessene Schätzgenauigkeit – die Terminologie

International wird zumeist von ‚accuracy‘ gesprochen, sollen Schätzergebnisse beurteilt werden (Swan & Jones, 1980; Siegel et al., 1982; Jones et al., 2012; Huang, 2014). Im deutschsprachigen Raum findet vor allem der Begriff der ‚Genauigkeit‘ Verwendung. Beide Begriffe überbetonen jedoch die ‚Welt der genauen Zahlen‘ mit ihrer Vorstellung von Präzision und Richtigkeit (vgl. Freudenthal, 1978). Der hier vorgeschlagene Terminus ‚angemessen‘ wirkt dagegen weniger objektiv und verweist sowohl auf die Situations- als auch die Beurteilungsabhängigkeit. Wird er darüber hinaus wörtlich genommen – gemessen mit Bezug auf ein Referenzmaß – dann stellt sich die Frage: Welche Referenzpunkte sind nützlich/sinnvoll, um die Angemessenheit eines Schätzergebnisses zu beurteilen?

Kriterien zur Beurteilung der Angemessenheit von Schätzergebnissen

In Ruwisch et al. (2015) wurden Beiträge zum Schätzen von Größen aus der fachdidaktischen Literatur im Detail mit Versuchspersonen, publizierten Items und Beurteilungsprozess vorgestellt. Zumeist werden Grenzen der prozentualen Abweichung vom Messergebnis als Beurteilungskriterien angegeben. Grob lässt sich zwischen dichotomer und polytomer Beurteilungspraxis unterscheiden (vgl. Tabelle 1).

In der psychologischen Forschung wird das Schätzen als komplexer kognitiver Vorgang genutzt, um abweichende kognitive Hirnstrukturen nachzuweisen. Schätzfragebogen werden i.d.R. eingesetzt, um kognitive Prozesse des Frontallappens zu messen (D’Aniello et al., 2015). Bereits durch die Zielsetzung wird deutlich, dass in diesem Rahmen weder die Art der zu

schätzenden Größe noch die normative Frage nach einer guten Schätzung interessiert. Die Testbatterien werden vielmehr anhand einer entsprechend großen Anzahl an Versuchspersonen ohne Beeinträchtigungen standardisiert und definieren “pathologische Schätzungen” mit Hilfe von Perzentilen oder Standardabweichungen vom Mittelwert dieser Versuchsgruppe (e.g. Axelrod & Millis, 1994; MacPherson et al., 2014). Auch in der kognitiven Psychologie findet sich eine gewisse Variabilität im Beurteilungsverhalten der Angemessenheit der Schätzwerte: dichotome oder polytome Beurteilung finden sich ebenso wie verschiedene Cut-off-Werte.

Studie	Versuchspersonen	Größenbereiche	Kriterium: Prozentuale Abweichung (d)
Dichotome Beurteilung			
Swan & Jones (1980)	4.–12. Klasse; College Studierende	Länge, Gewicht, Temperatur	$d < 25\%$
Hildreth (1980)	10-, 12- und 18-Jährige	Länge, Fläche	Anzahl der Aufgaben mit $d < 33\%$
Polytome Beurteilung			
Corle (1963)	Grundschullehrkräfte; College Studierende	Länge, Gewicht, Zeit, Volumen, Temperatur	$d < 10\%$ $10\% \leq d \leq 100\%$ $d > 100\%$
Siegel et al. (1982)	2.–8. Klasse	Länge (Anzahl)	accurate: $< 50\%$ reasonable: innerhalb einer Größenordnung
Hogan & Brezinski (2003)	College Studierende	Länge, Gewicht, Volumen, Zeit	3 Punkte: $\leq 10\%$ 2 Punkte: $10\% < d \leq 20\%$ 1 Punkt: $20\% < d \leq 30\%$
Huang (2014)	4.–6. Klasse	Länge, Fläche	accurate – 2 Punkte: $d \leq 10\%$ acceptable – 1 Punkt: $10\% < d \leq 25\%$

Tab. 1: Kriterien zur Beurteilung von Schätzergebnissen

Ergebnisse der Anwendung verschiedener Kriterien in zwei Studien zum Schätzen von Grundschulkindern

In einer Interviewstudie von Heid (2017) mit 250 ViertklässlerInnen zu ihren Schätzstrategien im Bereich Längen und Fassungsvermögen und einer 29 Aufgaben umfassenden eigenen Paper-and-Pencil-Studie mit 187 Dritt- und ViertklässlerInnen zum Schätzen von Längen zeigte sich,

- dass Längen im Durchschnitt statistisch signifikant besser geschätzt werden als das Fassungsvermögen;
- dass in beiden Größenbereichen eher unter- als überschätzt wird (im Durchschnitt wiesen die schriftlichen Schätzwerte eine 10%-ige Unterschätzung auf);
- dass in allen Teilbereichen ähnlich viele Antworten eine max. 10%-ige Abweichung vom Realwert aufweisen (18% bei den Längen; 15 % beim Fassungsvermögen; 15 % im schriftlichen Längenschätztest), dass statistisch signifikante Unterschiede in den anderen Intervallen zu finden sind: Während 77 % der geschätzten Längen innerhalb des $\pm 50\%$ -Intervalls lagen, waren es nur 44 % der geschätzten Volumina. Im Längenschätztest lagen 45 % im $\pm 30\%$ -Intervall und 66 % im $\pm 50\%$ -Intervall.
- dass unterschiedliche Beurteilungsschemata statistisch signifikante Unterschiede in der Schätzfähigkeit ergeben (s. Abb. 1);

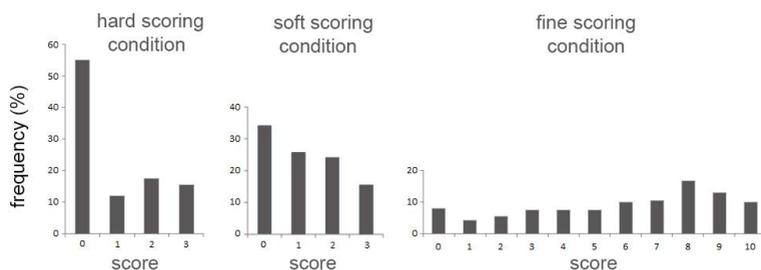


Abb. 1: Ergebnisse beim Beurteilen mit harten (10%, 20%, 30%, >30%), weichen (10%, 25%, 50%, >50%) und feinen (jeweils 10%-Schritte) Grenzen

Diskussion und Ausblick

Als grundsätzliche Frage steht im Raum, ob es sich bei der Beurteilung von Schätzergebnissen um eine normative – wie in der Mathematikdidaktik unterstellt – oder eine empirische – wie in der psychologischen Forschung genutzt – Frage handeln soll.

Wenn normativ entschieden werden soll, wann ein hinreichendes, gutes oder gar nicht genügendes Schätzergebnis vorliegt, so scheint allgemein das obige polytome ‚Soft-Scoring‘ am ehesten geeignet, zwischen verschiedenen Gruppen von Schätzergebnissen zu differenzieren.

Ungeklärt ist jedoch, wovon die Wahl der Grenzen abhängen sollte: Wenn die Schätzungen zum Fassungsvermögen bei Kindern wie ExpertInnen schlechter ausfallen, sollten sie dann auch weiter gesetzt werden? Wie ist die Größenordnung der jeweiligen Items zu berücksichtigen?

Und schließlich: Sind die Grenzen altersabhängig zu wählen?

Literatur

- Axelrod, B. N. & Millis, S. R. (1994). Preliminary standardization of the cognitive estimation test. *Psychological Assessment*, 1(3), 269–274.
- D’Aniello, G.E., Castelnuovo, G. & Scarpina, F. (2015). Could cognitive estimation ability be a measure of cognitive reserve? *Frontiers in Psychology*, 6, 1–4.
- Corle, C.G. (1963). Estimates of quantity by elementary teachers and college juniors. *The Arithmetic Teacher*, 10(6), 347–353.
- Forrester, M.A., Latham, J. & Shire, B. (1990). Exploring estimation in young school children. *Educational Psychology*, 10(4), 283–300.
- Freudenthal, H. (1978). *Vorrede zu einer Wissenschaft vom Mathematikunterricht*. München: Oldenbourg.
- Heid, L.M. (2017). *Das Schätzen von Längen und Fassungsvermögen. Eine Interviewstudie zu Strategien mit Kindern im 4. Schuljahr*. Wiesbaden: Springer.
- Hildreth, D. J. (1980). *Estimation Strategy Use in Length and Area Measurement Tasks by Fifth and Seventh Grade Students*. [http://phdtree.org/24304583-estimation-strategy-use-in-length-and-area-measurement-tasks-by-fifth-and-seventh-grade-students/\[22-09-14\]](http://phdtree.org/24304583-estimation-strategy-use-in-length-and-area-measurement-tasks-by-fifth-and-seventh-grade-students/[22-09-14]).
- Hogan, T. P. & Brezinski, K. L. (2003). Quantitative Estimation: One, Two, or Three Abilities? *Mathematical Thinking and Learning*, 5(4), 259–280.
- Huang, H.-M. E. (2014). Investigating Children’s Ability to Solve Measurement Estimation Problems. In *PME 38*, Vol. 3, (353–360). Vancouver: PME.
- Jones, M. G.; Forrester, J. H.; Gardner, G. E.; Andre, T. & Taylor, A. R. (2012). Students’ Accuracy of Measurement Estimation: Context, Units, and Logical Thinking. *School Science and Mathematics*, 112(3), 171–178.
- MacPherson, S. E., Wagner, G. P., Murphy, P., Bozzali, M., Cipolotti, L. & Shallice, T. (2014). Bringing the cognitive estimation test into the 21st century: Normative data on two new parallel forms. *PLoS One*, 9(3), e92554. doi: 10.1371/journal.pone.0092554.
- Ruwisch, S., Heid, M. & Weiher, D. F. (2015). Measurement estimation in primary school: Which answer is adequate? *PME 39*, 4, 113–120. Hobart: PME.
- Siegel, A. W.; Goldsmith, L. T. & Madson, C. R. (1982). Skill in Estimation Problems of Extent and Numerosity. *Journal for Research in Mathematics Education*, 13(3), 211–232.
- Swan, M. & Jones, O. E. (1980). Comparison of Student’s Percepts of Distance, Weight, Height, Area, and Temperature. *Science Education*, 64(3), 297–307.