

Scheer, D. (2022). Wie oft oder wie verzögert tritt ein Verhalten auf?. Time-Sampling-Techniken in der Verhaltensbeobachtung. In M. Gebhardt, D. Scheer & M. Schurig (Hrsg.), *Handbuch der sonderpädagogischen Diagnostik. Grundlagen und Konzepte der Statusdiagnostik, Prozessdiagnostik und Förderplanung* (S. 649-656). Regensburg: Universitätsbibliothek. <https://doi.org/10.5283/epub.53149>

Wie oft oder wie verzögert tritt ein Verhalten auf?

Time-Sampling-Techniken in der Verhaltensbeobachtung

David Scheer

Sowohl in der Diagnostik als auch in der Forschung interessiert uns oftmals, wie häufig ein*e Schüler*in ein bestimmtes Verhalten zeigt oder wie lange es dauert, bis ein*e Schüler*in auf ein Ereignis (z. B. Arbeitsanweisung) reagiert.

In der Praxis werden dafür manchmal einfache Strichlisten (Häufigkeit) oder der Blick auf die Uhr (Verzögerung) genutzt. Das bringt methodisch einige Probleme und Unschärfen mit sich. Für Strichlisten zeigt Scheer (2021) die Problematik am Beispiel des Verhaltens »Schimpfwörter benutzen« auf: Ist ein*e Schüler*in in einen Streit verwickelt, so stellt sich die Frage, ob jetzt jedes Schimpfwort einzeln gezählt werden soll, oder nur die gesamte Sequenz. Ein anderes Beispiel wäre »Konzentriertes Arbeiten am Platz«: Unterbricht ein*e Schüler*in die Arbeit kurz und setzt sie dann fort, so würde auf der Strichliste 2x das Zielverhalten notiert. Das heißt: Diese Form der Häufigkeitsmessung ist ungenau und nicht besonders reliabel.

Abhilfe schaffen Time-Sampling-Techniken (Salvia et al., 2010; Scheer, 2021; Schmidt-Atzert & Amelang, 2012). Eine detaillierte und praxisorientierte Darstellung der systematischen Verhaltensbeobachtung findet sich bei Scheer (2021), die Time-Sampling-Strategien werden ausführlich bei Salvia et al. (2010) ausgeführt. In diesem Beitrag geht es vor allem um die Praxis des Time-Samplings, die Durchführung mithilfe der kostenfreien App MySO (<https://www.studentobservation.com/>), verfügbar für Android und iOS) sowie einer Zusammenfassung des Forschungsstands zur Zuverlässigkeit von Time-Sampling.

1 Wie funktioniert Time-Sampling?

Die theoretische Hauptannahme hinter Time-Sampling-Techniken ist die, dass in der direkten Verhaltensbeobachtung ein kontinuierlicher Strom von Verhalten abläuft, der von einer*inem Beobachter*in sequenziert und segmentiert werden muss (Schmidt-Atzert & Amelang, 2012):

»Bei einer systematischen Verhaltensbeobachtung wird keine vollständige Beschreibung des Verhaltens angestrebt, sondern es soll immer nur ein bestimmter Teilaspekt des Verhaltens erfasst werden, beispielsweise Aggressivität, Kooperationsverhalten großer Aufwand Selektion, Segmentierung, und Quantifizierung

standardisieren oder Durchsetzungsfähigkeit. Die Vielfalt der Verhaltensweisen, die einen solchen Teilbereich des Verhaltens ausmachen, soll auf wenige Aussagen reduziert und zudem quantifiziert werden.« (Schmidt-Atzert & Amelang, 2012, S. 314–315)

Beim Time-Sampling wird dieser »Strom«, also der Beobachtungszeitraum, in (kurze) Intervalle (z. B. 10 Sekunden, 15 Sekunden oder 30 Sekunden) unterteilt. Mithilfe einer Stoppuhr oder eines Timers wird dann während der Beobachtung für jedes einzelne Intervall beurteilt, ob das zu beobachtende Zielverhalten gezeigt wurde oder nicht (Salvia et al., 2010; Schmidt-Atzert & Amelang, 2012). Schmidt-Atzert & Amelang (2012) verweisen auf die Notwendigkeit, dass die*der Beobachter*in über Kopfhörer jeweils zum Ende des Intervalls ein akustisches Signal bekommt – mittlerweile erlauben Smartphones auch einen stillen Hinweis über Vibration.

So kann, unabhängig von der beobachtenden Person die Häufigkeit einzelner Verhaltensweisen ermittelt werden. Markiert man in den Intervallen auch noch, ob eine bestimmte Intervention oder Arbeitsanweisung der Lehrkraft gegeben wurde, kann damit auch erfasst werden, wie bzw. mit welcher zeitlichen Verzögerung ein*e Schüler*in darauf reagiert.

Generell eignet sich das Time-Sampling sowohl für statusdiagnostische Fragestellungen (dann sollten aber mehrere Verhaltensproben genommen werden), z. B. um Prioritäten bei Förderzielen festzulegen, als auch für prozessdiagnostische Fragen, z. B. zur Evaluation von Fördermaßnahmen.

2 Welche Formen des Time-Samplings gibt es?

Es gibt drei Formen des Time-Samplings:

- *Whole-Interval-recording (WIR)*: Ein Intervall wird dann markiert, wenn das Zielverhalten (z. B. »Konzentriertes Arbeiten am Platz«) von Beginn bis Ende des Intervalls ohne Unterbrechung gezeigt wurde.
- *Partial-Interval-Recording (PIR)*: Ein Intervall wird dann markiert, wenn das Zielverhalten zur irgendeinem Zeitpunkt des Intervalls gezeigt wurde.
- *Momentary-Time-Sampling (MTS)*: Es wird markiert, ob das Zielverhalten im Moment des Intervallwechsels (beim Timer-Signal) gezeigt wurde.

3 Wie gehe ich vor?

- Im ersten Schritt muss ein Zielverhalten definiert werden, das ich bei der*dem Schüler*in beobachten möchte. Es kann sich dabei sowohl um erwünschtes Verhalten als auch um unerwünschtes Verhalten handeln. Als allgemeine Beispiele kann man sich an den Items bekannter screening- oder verlaufsdiagnostischer Instrumente orientieren, z. B. dem DBR-PUTSIE (Schurig et al., 2020).
- Dann muss das Beobachtungs-Setting festgelegt werden. Da man *immer* mehrere Messzeitpunkte benötigt, sollten identische Bedingungen für alle Messzeitpunkte festgelegt werden. Möchte man verschiedene Kontexte vergleichen (z. B. die Frage: »Verhält sich

das Kind in Mathe anders als in Sachkunde?«), dann muss man für beide Kontexte separat jeweils mehrere Messungen unter je identischen Bedingungen vornehmen.

- Es muss festgelegt werden, wie lang jeder Beobachtungszeitraum sein soll (z. B. 15 Minuten). Alle Beobachtungszeiträume müssen gleich lang sein.
- Es muss festgelegt werden, wie lang die Intervalle für das Time-Sampling sein sollen (z. B. 10 Sekunden).
- Dann kann die Beobachtung durchgeführt werden. Das Ergebnis kann in ein Verlaufsdiagramm (Liniendiagramm) eingetragen werden. So kann ich zum Beispiel prüfen, ob eine pädagogische Unterstützungsmaßnahme die erhoffte Wirkung zeigt.

4 Wie benutze ich die App »MySO«?

Die App »MyOS« (<<https://www.studentobservation.com//>>) ist für Android und iOS kostenfrei in den jeweiligen App-Stores erhältlich. Sie dient der Smartphone-gestützten Durchführung von Schüler*innenbeobachtungen und kann gut für Time-Sampling-Prozeduren genutzt werden, am einfachsten für MTS. Die Oberfläche der App ist auf Englisch, was aber kein Problem darstellen sollte. Es ist aber darauf hinzuweisen, dass es sich um proprietäre Software eines kommerziellen Anbieters handelt, sodass hier der Anonymisierung/Pseudonymisierung von Schüler*innendaten vor der Erhebung eine besondere Bedeutung zukommt.

4.1 Einstellungen

In den Einstellungen wird zunächst folgendes festgelegt:

- *Engagement Codes*: Hier werden mit einem Kürzel und einer kurzen Beschreibung beobachtbare Verhaltensweisen festgelegt. Die App sieht zwar ausschließlich akademische Mitarbeit im Unterricht (»Engagement«) als Zielverhalten an, daran sind User jedoch nicht gebunden.
- *Intervention Codes*: Zusätzlich zum Schüler*innenverhalten können auch Lehrkräfteverhalten und/oder pädagogische Interventionen definiert werden, die mitbeobachtet werden. So kann beispielsweise erfasst werden, auf welche Art der Ansprache ein*e Schüler*in besonders reagiert. Unter der Rubrik Intervention Codes könnten aber auch andere vorab beschriebene im Klassenraum mögliche Ereignisse codiert und erfasst werden. Es ist aber darauf zu achten, dass (genau wie bei den Engagement Codes) zu viele unterschiedliche Verhaltens- oder Ereigniskategorien die Beobachtung komplexer machen würden.
- *Add/Remove Students*: Schüler*innen müssen vor der Beobachtung in der App angelegt werden. Es werden Name (bitte pseudonymisieren) und Geburtsdatum (bitte fiktiv angeben) erfasst.
- *Add/Remove Class*: Hier werden die Unterrichtsstunden/Klassen, in denen Beobachtungen durchgeführt werden sollen, angelegt. Diese werden mit Name (z. B. »Mathematik, Klasse 8a, donnerstags«), Uhrzeit und (optionaler) Beschreibung eingetragen.

4.2 Beobachtung

Sind alle erforderlichen Einstellungen vorgenommen, kann mit dem Button »Observe« eine neue Beobachtungssession gestartet werden. Vor Beginn der Beobachtung werden Schüler*in und Unterrichtsstunde/Klasse ausgewählt sowie die Methode, mit der die*der Beobachter*in über die Intervalle informiert wird (Vibration oder Ton). Der Beobachtungszeitraum kann in Fünfminutenschritten von fünf Minuten bis zwei Stunden eingestellt werden, die Intervalllänge in 15-Sekundenschritten von 15 Sekunden bis fünf Minuten. Auf der nächsten Seite werden Notizen, etwa zum Grund der Beobachtung erfasst. Auch kann angegeben werden, welche*r (unauffällige) Schüler*in als Vergleichsperson mitbeobachtet wird.

Während der Beobachtung wird der*die User*in mit Ablauf eines Intervalls durch Vibration oder Ton informiert und gebeten, einen Engagement- und/oder Intervention-Code auszuwählen.

4.3 Berichte

Die App ermöglicht es, die Ergebnisse der einzelnen Beobachtungssessions als HTML-Datei entweder per Mail oder über andere Apps zu exportieren. Ein CSV- oder Excel-Export ist leider nicht möglich.

5 Ist Time-Sampling zuverlässig und gültig?

In der empirischen Forschung zeigt sich, dass die Durchführungsform MTS als valide und reliabel angesehen werden kann, während die Durchführungsmethoden WIR und PIR eher ungenau zu sein scheinen. Das ist aber situationsabhängig. MTS bietet sich bei Verhaltensweisen mit etwas längerer Dauer (z. B. Mitarbeit, Arbeitsverweigerung,...) an, während sich PIR eher bei punktuellen Ereignissen (z. B. Hereinrufen) eignet.

Salvia et al. (2010) verweisen auf ältere Befunde, dass die Durchführungsweisen WIR und PIR zu eher ungenauen Schätzungen der Dauer und Häufigkeit von Verhalten führen, während MTS bei kurzen Intervallen (10- und 15-Sekunden) zu akkuraten Schätzungen führe. Auch Fiske & Delmolino (2012) betonen die generelle Überlegenheit von MTS. Allerdings konstatieren sie, dass MTS die Häufigkeit bei Verhalten mit geringer Auftretensrate unterschätzen kann und in diesen Fällen gegenüber geringen Veränderungen nicht hinreichend sensitiv ist.

Dem hingegen ist die aktuelle empirische Forschungslage sehr heterogen und muss differenziert betrachtet werden. Dies hängt insbesondere mit der Heterogenität der untersuchten Kontexte, Verhaltensdimensionen und Schülerpopulationen zusammen.

Becraft et al. (2016) konnten für das Verhalten »Sprechen« bei Vorschulkindern hohe Werte an Interval-by-Interval Interobserver-Agreement bei MTS feststellen, d.h. die Übereinstimmung der Beobachter auf Ebene der Einzelintervalle war hoch. Zudem ergab die Verwendung von 10-, 20-, 30-, 60- und 90-Sekunden-Intervallen in dieser Studie weniger als 10% Abweichung von der Beobachtung mittels 5-Sekunden-Intervallen. In Bezug auf aktive Mitarbeit im Unterricht erweisen sich beim MTS bessere Werte hinsichtlich der Zuverlässigkeit, wenn kürzere (10 oder 15 Sekunden) statt längerer (20 oder 30 Sekunden) Intervalle genutzt werden, wobei eine gute Zuverlässigkeit im Sinne der Generalisierungstheorie nur bei mindestens zwei Beobachtungszeitpunkten erreicht werden (Briesch et al., 2017). Insgesamt ist die Übereinstimmung von MTS-

Daten mit kontinuierlichen Messungen umso höher, je kürzer die Intervalle sind (Zakszeski et al., 2017). Zudem sind, in Bezug auf aktive Mitarbeit im Unterricht, die Ergebnisse exakter bei Schüler*innen mit hoher als mit niedriger aktiver Unterrichtsbeteiligung. Auch zeigen Briesch et al. (2017) für problematisches Verhalten, dass bei Schüler*innen mit erhöhtem Risiko doppelt so viele Beobachtungszeitpunkte nötig sind wie einer Normalstichprobe, um reliable Ergebnisse zu bekommen. Im Vergleich der verschiedenen Time-Samplings zeigt sich, dass für eine hohe Zuverlässigkeit im Sinne der Generalisierungstheorie beim MTS 1-2 Beobachter genügen, während WIR 2-3 Beobachter und PIR gar 3-7 Beobachter benötigt, deren Ergebnisse gemittelt werden (Johnson et al., 2017).

Beim PIR zeigt sich in einer Simulationsstudie, dass kürzere Intervalle und eine niedrigere Auftretensrate des Verhaltens zu geringeren Abweichungen der Ergebnisse von der tatsächlichen Auftretenshäufigkeit führen und dass PIR insbesondere dann geeignet scheint, wenn Verhaltensweisen ca. 1 Sekunde dauern (Yoder et al., 2018).

Gleichzeitig kommen Lane & Ledford (2014) zu dem Schluss, dass alle drei Methoden ungeeignet zur Erfassung der Verhaltenshäufigkeit seien, MTS jedoch recht Akkurat die Dauer von Verhalten messen könne. Time-Sampling korrespondiert hier bei Intervalllängen von höchstens 3 Minuten am stärksten mit kontinuierlichen Messungen, wobei MTS akkurater ausfällt als PIR (LeBlanc et al., 2020). Im Gegensatz dazu erweisen sich sowohl MTS als auch PIR bei Schüler*innen mit Autismus-Spektrum-Störungen als adäquate Maße für die Messung von Verhaltenshäufigkeit, wobei MTS hier seltene Verhaltensweisen unterschätzte (Sam et al., 2015). Eine Überlegenheit von MTS über PIR, aber auch über WIR in Bezug auf die Messung von Verhaltensdauer zeigt sich bei Ledford et al. (2015), allerdings erhöht sich bei MTS die Gefahr von Typ-II-Fehlern in Interventionsstudien, wenn besonders kurze Beobachtungssessions genutzt werden. Auch die Untersuchung von Wood et al. (2016) zeigt für alle Time-Sampling-Methoden eine hohe Annäherung an tatsächliche kontinuierliche Verhaltensmessungen mit einer leichten Überlegenheit von MTS. Allerdings erwies sich der Unterschied zwischen MTS und PIR hinsichtlich der Genauigkeit als geringer als in früherer Literatur angenommen.

6 Einsatzmöglichkeiten und Grenzen des Time-Samplings

Insgesamt scheint sich MTS besonders gut bei Verhaltensweisen zu eignen, die nicht nur punktuell, sondern mit einer gewissen Dauer auftreten (z. B. Mitarbeit, Arbeitsverweigerung, Streitigkeiten, aggressives Verhalten etc.), sollte aber mit eher kurzen Intervallen (10 Sekunden oder 15 Sekunden) genutzt werden. Sehr punktuell auftretende Verhaltensweisen (z. B. Reinrufen im Unterricht) hingegen sollten eher traditionell mittels Strichlisten erfasst werden und benötigen keine genauere Erfassung mittels Time-Sampling.

Auch, wenn MTS dank moderner Softwarelösungen (und auch der Möglichkeit von Videografie) mittlerweile ökonomischer umsetzbar ist als früher, ist die Durchführungsökonomie doch nach wie vor schlechter als bei der ebenfalls sehr reliablen direkten Verhaltensbeurteilung (Casale et al., 2019). Insbesondere bei sehr wichtigen diagnostischen Entscheidungen (z. B. bei potenziellen Schulwechsellern) sollte die Beobachtung mittels MTS von 2-3 geschulten Beobachtern vorgenommen werden, was sehr zeit- und damit auch kostenintensiv ist.

Literatur

- Becraft, J. L., Borrero, J. C., Davis, B. J., & Mendres-Smith, A. E. (2016). Assessment of a rotating time sampling procedure: Implications for interobserver agreement and response measurement. *Education & Treatment of Children, 39*(1), 1–20.
- Briesch, A. M., Ferguson, T. D., Daniels, B., Volpe, R. J., & Feinberg, A. B. (2017). Examining the influence of interval length on the dependability of observational estimates. *School Psychology Review, 46*(4), 426–432. <<https://doi.org/10.17105/SPR-2016-0006.V46-4/>>
- Casale, G., Huber, C., Hennemann, T., & Grosche, M. (2019). *Direkte verhaltensbeurteilung in der schule. Eine einföhrung für die praxis*. Reinhardt.
- Fiske, K., & Delmolino, L. (2012). Use of discontinuous methods of data collection in behavioral intervention: Guidelines for practitioners. *Behavior Analysis in Practice, 5*(2), 77–81. <<https://doi.org/10.1007/BF03391826/>>
- Johnson, A. H., Chafouleas, S. M., & Briesch, A. M. (2017). Dependability of Data Derived from Time Sampling Methods with Multiple Observation Targets. *School Psychology Quarterly, 32*(1), 22–34. <<https://doi.org/10.1037/spq0000159/>>
- Lane, J. D., & Ledford, J. R. (2014). Using Interval-Based Systems to Measure Behavior in Early Childhood Special Education and Early Intervention. *Topics in Early Childhood Special Education, 34*(2), 83–93. <<https://doi.org/10.1177/0271121414524063/>>
- LeBlanc, L. A., Lund, C., Kookan, C., Lund, J. B., & Fisher, W. W. (2020). Procedures and accuracy of discontinuous measurement of problem behavior in common practice of applied behavior analysis. *Behavior Analysis in Practice, 13*(2), 411–420. <<https://doi.org/10.1007/s40617-019-00361-6/>>
- Ledford, J. R., Ayres, K. M., Lane, J. D., & Lam, M. F. (2015). Identifying Issues and Concerns with the Use of Interval-Based Systems in Single Case Research Using a Pilot Simulation Study. *Journal of Special Education, 49*(2), 104–117. <<https://doi.org/10.1177/0022466915568975/>>
- Salvia, J., Ysseldyke, J. E., & Bolt, S. (2010). *Assessment in special and inclusive education* (11th ed.). Wadsworth/Cengage Learning.
- Sam, A., Reszka, S., Odom, S., Hume, K., & Boyd, B. (2015). The Use of Coding Methods to Estimate the Social Behavior Directed toward Peers and Adults of Preschoolers with ASD in TEACCH, LEAP, and Eclectic »BAU« Classrooms. *Behavioral Disorders, 40*(2), 91–101.
- Scheer, D. (2021). *Toolbox Diagnostik*. Kohlhammer.
- Schmidt-Atzert, L., & Amelang, M. (2012). *Psychologische Diagnostik* (5th ed.). Springer. <<https://doi.org/10.1007/978-3-642-17001-0/>>
- Schurig, M., Garthe-Krause, S., & Gebhardt, M. (2020). *Verlaufsdagnostik für Verhalten in der Schule (DBR-PUTSIE)*. Unpublished. <<https://doi.org/10.13140/RG.2.2.33359.84641/>>
- Wood, B. K., Hojnoski, R. L., Laracy, S. D., & Olson, C. L. (2016). Comparison of Observational Methods and Their Relation to Ratings of Engagement in Young Children. *Topics in Early*

Childhood Special Education, 35(4), 211–222. <<https://doi.org/10.1177/0271121414565911>>

Yoder, P. J., Ledford, J. R., Harbison, A. L., & Tapp, J. T. (2018). Partial-Interval Estimation of Count: Uncorrected and Poisson-Corrected Error Levels. *Journal of Early Intervention*, 40(1), 39–51. <<https://doi.org/10.1177/1053815117748407>>

Zakszeski, B. N., Hojnoski, R. L., & Wood, B. K. (2017). Considerations for Time Sampling Interval Durations in the Measurement of Young Children’s Classroom Engagement. *Topics in Early Childhood Special Education*, 37(1), 42–53. <<https://doi.org/10.1177/0271121416659054>>

Prof. Dr. David Scheer ist Professor für Pädagogik und Didaktik im sonderpädagogischen Förderungsschwerpunkt Lernen an der Pädagogischen Hochschule Ludwigsburg. Die Schwerpunkte seiner Forschung liegen auf inklusionsorientierten Schulentwicklungsprozessen, auf pädagogischen Unterstützungsmaßnahmen für Schüler*innen mit sonderpädagogischem Unterstützungsbedarf des Lernens und der sozial-emotionalen Entwicklung sowie auf Fragen der inklusiven Lehrkräftebildung. <https://orcid.org/0000-0002-0534-7869>

