

DLG-Expertenwissen 2/2020

Praxisleitfaden zur Panelschulung

Teil 3: Messung der Leistungsfähigkeit sensorischer Prüfer und Panels



1. Einleitung und Hintergrund

In den Teilen 1 und 2 des Praxisleitfadens zur Panelschulung wurden die Anforderungen nach DIN EN ISO 8586:2014-05 und Schulungspläne zur Prüfer- und Panelqualifizierung vorgestellt. Im Teil 3 steht nun die Messung der Leistungsfähigkeit sensorischer Einzelprüfer sowie ganzer Panels mit methodisch unterschiedlicher Ausrichtung im Fokus. Der Leitfaden möchte auf die wesentlichen zu beachtenden Aspekte aufmerksam machen, die bei der Messung der Leistungsfähigkeit von Prüfpersonen und Panels zu berücksichtigen sind. Es werden zunächst Hintergründe und Kriterien zur Bestimmung der Leistungsfähigkeit dargelegt sowie ausgewählte Methoden zu ihrer Messung und Optimierung vorgestellt, welche auch ohne fortgeschrittene statistische Kenntnisse durchführbar sind. Fallbeispiele sollen verschiedene Anwendungen praxisnah verdeutlichen, so dass im Bereich Lebensmittelsensorik tätige Fach- und Führungskräfte wertvolle Anregungen für die praktische Umsetzung erhalten, die jedoch stets an die jeweilige Fragestellung im eigenen Unternehmen und an die dort durchgeführten Sensorikprojekte angepasst werden müssen.

Eine wesentliche Voraussetzung für verlässliche Prüf- und Panelergebnisse sowie für valide Daten in der Humansensorik stellt neben dem Vorhandensein einer für sensorische Prüfungen maßgebenden technischen Infrastruktur und der Einhaltung der „Guten Sensorikpraxis“, die Qualifikation des sensorischen Personals dar. Die hohe fachliche Bedeutung eines leistungsfähigen Panels, welches in der Lebensmittelwirtschaft die Produktentwicklung sowie die Qualitätssicherung auch im Sinne von Food Fraud und Lebensmittelsicherheit (vgl. IFS, BRC-Standard u. a.) ideal unterstützt, wird nicht zuletzt auch durch eine schnelle Amortisierung von in die Lebensmittelsensorik investiertem Kapital deutlich (vgl. DLG Trendmonitor 2019).

Betroffen vom Nachweis fachgerecht qualifizierter und in der Leistungsfähigkeit überwachter Prüfer und Panels sind zudem auch alle Organisationen, die als Prüfinstitut oder Prüflabor für Dritte tätig sind, in ihrer Lebensmittelanalytik sensorische Methoden einsetzen und nach ISO/IEC 17025:2018-03 akkreditiert wurden.

2. Bedeutung des Themas

Im Laborbetrieb und im Unternehmensalltag ist es notwendig, auf die Zuverlässigkeit der verwendeten Messeinrichtungen vertrauen zu können. So ist vor der Anwendung zu prüfen, ob die Messgeräte für den jeweiligen Einsatzzweck geeignet sind, den Messbereich abdecken und die erforderliche Messgenauigkeit besitzen. Während des praktischen Betriebs ist die Leistungsfähigkeit der Messgeräte in vorgegebenen oder vereinbarten Intervallen zu prüfen, um mögliche Abweichungen zu erkennen und entsprechende Maßnahmen ergreifen zu können. Analog zu technischen Messeinrichtungen kann auch eine Prüfergruppe bzw. ein Panel als analytisches Messinstrument betrachtet werden, das ähnliche Vorgaben zu erfüllen hat.

Prinzipiell erwartet man von einem Messgerät, dass es richtige und präzise Ergebnisse liefert. **Richtige Ergebnisse** stimmen mit dem wahren Wert überein oder zeigen nur eine geringe Abweichung davon (vgl. DIN ISO 5725-2:2012). In der chemischen Analytik kann man vergleichsweise einfach von einem wahren Wert ausgehen, wenn man eine Probe untersucht, in der z. B. eine bestimmte Kochsalzmenge in entmineralisiertem Wasser gelöst worden ist. Da in der Lebensmittelsensorik jedoch die menschliche Wahrnehmung und die Einordnung eines Reizes zu berücksichtigen sind, wird gerade bei üblichen Lebensmitteln, die meist komplexe Gemische sind, der Mittelwert aus mehreren Messergebnissen einer Prüfergruppe bzw. eines Panels als Ersatz für den wahren Wert einer Merkmalseigenschaft genommen. Der wahre Wert in der Lebensmittelsensorik entspricht somit einer relativen Messung in Abhängigkeit von der Lebensmittelmatrix und der Interaktion mit anderen Sinnesmodalitäten.

Als **Wiederholbarkeit oder Präzision** wird die Fähigkeit beschrieben, bei mehrmaliger Untersuchung derselben Probe übereinstimmende Messergebnisse zu erzielen. Die Präzision darf dabei nicht mit der Richtigkeit verwechselt werden (vgl. DIN ISO 5725-2:2012). Ein Messgerät kann durchaus wiederholt übereinstimmende Ergebnisse liefern und somit eine hohe Präzision zeigen. Wenn aber diese präzisen Wiederholungen vom wahren Wert weit entfernt sind, ist die Richtigkeit des Ergebnisses nicht genügend. Um dies zu verdeutlichen wird nachfolgend zunächst eine Übersicht über die jeweiligen

fachlichen Anforderungen an Prüfpersonen gegeben, bevor dann die Kriterien zur Prüfung und zum Monitoring der Leistungsfähigkeit dargelegt werden.

Die detaillierten fachlichen Anforderungen an Prüfpersonen und Panels variieren in der Praxis in der Regel im Hinblick auf ihr jeweiliges spezifisches Einsatzgebiet, so dass das regelmäßige Training und auch das Monitoring darauf fokussiert sein sollten.

Während das Screening zur Gewinnung sensorisch-analytischer Prüfer eher grundlegend die allgemeinen sensorischen Fähigkeiten der rekrutierten Prüfpersonen testet und u. a. das Farbsehvermögen (ggf. Farbblindheit) oder auch Anosmien ermittelt, Geruchs- und Geschmacksschwellen erfasst, die Sensitivität bei der Erkennung von Unterschieden im Geruch, Geschmack oder der Textur überprüft und die generelle sensorische Sprach- und Ausdrucksfähigkeit beim Beschreiben von sensorischen Empfindungen thematisiert, ist die sich an die Prüferselektion anschließende vertiefende Schulungs- und Qualifizierungsphase stärker spezialisiert und an den späteren Einsatzbereichen der Prüfer ausgerichtet.

Wie in den DLG-Expertenwissen Teil 1 (7/2017) und Teil 2 (12/2018) beschrieben, sind auf Basis der jeweiligen Einsatzbereiche entsprechende Anforderungsprofile an die Prüfer und an ihre sensorische Leistungsfähigkeit zu definieren. Darauf aufbauend sind sodann spezielle Schulungspläne und sensorische Prüfungen zu entwickeln sowie zu erreichende Mindestanforderungen zu formulieren, die eine gezielte Auswahl und weitere Selektion von Prüfpersonen aufgrund ihrer sensorischen Wahrnehmungen und Messkompetenzen ermöglichen. Hingewiesen sei in diesem Kontext zudem nochmals darauf, dass man, um letztlich die zur statistischen Absicherung sensorisch-analytischer Ergebnisse erforderliche Panelgröße zu erhalten, im Rahmen des Screenings mindestens (wenn möglich) die doppelte bis dreifache Anzahl möglicher Prüfpersonen einbinden sollte.

In Abbildung 1 sind die gängigsten Applikationen sensorischer Prüfer bzw. Panels dargelegt. Wie dort dargestellt, unterscheidet man bei den sensorisch-analytischen Methoden bezüglich der jeweiligen Problemstellung generell die beiden Verfahrensgruppen Unterschiedsprüfungen (diskriminierende Prüfungen) und beschreibende Prüfungen (deskriptive Prüfungen). Darüber hinaus gibt es Methoden, die zwar den diskriminierenden Prüfungen zugeordnet werden, aber in Form von Varianten Elemente aus dem Bereich der deskriptiven Verfahren nutzen:

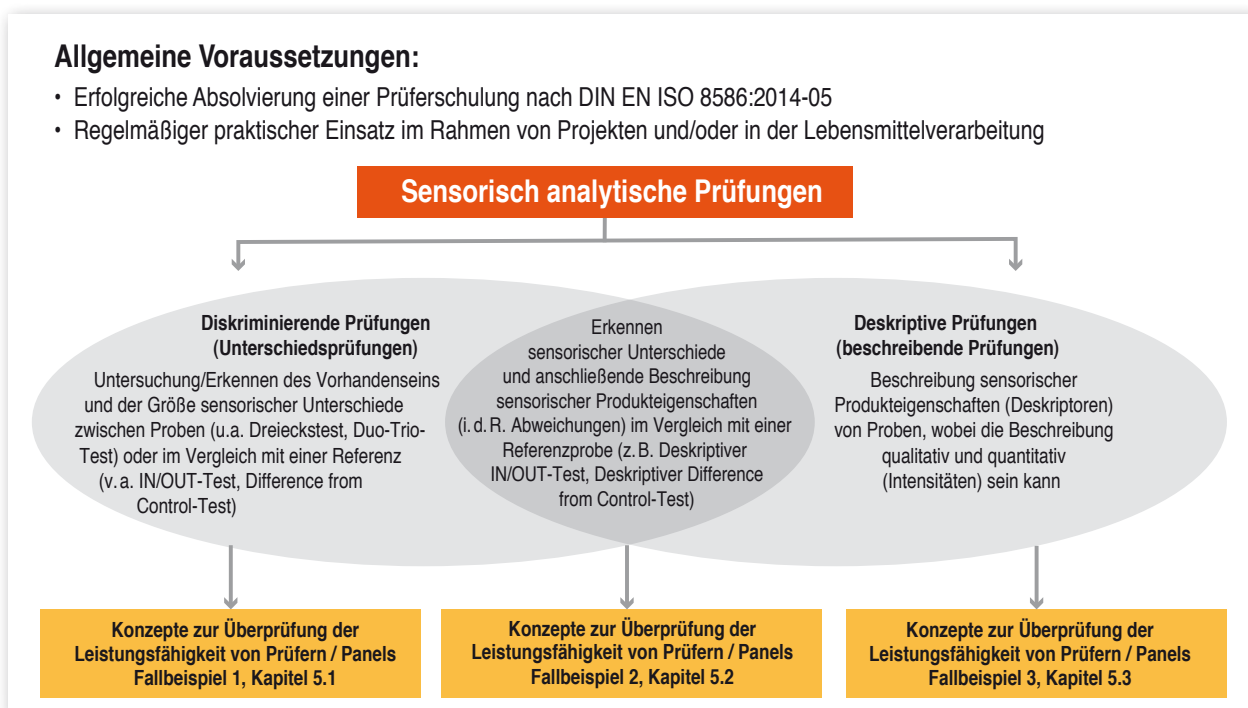


Abbildung 1: Einsatzbereiche sensorisch-analytischer Prüfer und Panels

- **Diskriminierende Prüfungen**

- Produktvergleich zur Ermittlung sensorisch erfassbarer (geringer) Unterschiede (z. B. Dreiecksprüfung, Duo-Trio-Prüfung, Paarweise Vergleichsprüfung, Rangordnungsprüfung);
- Produktvergleich v. a. im Rahmen der sensorischen Qualitätskontrolle (z. B. IN/OUT-Test, Difference from Control-Test unter Einbeziehung von Referenzmustern)

Diese sind methodisch primär Unterschiedsprüfungen, die jedoch in der jeweiligen deskriptiven Variante (deskriptiver IN/OUT-Test bzw. deskriptiver Difference from Control-Test) um beschreibende Elemente aus den deskriptiven Prüfungsverfahren erweitert werden können.

- **Deskriptive Prüfungen**

Ermittlung und Beschreibung sensorischer Produkteigenschaften; zusätzlich kann bei einzelnen Verfahren auch die Intensitätsausprägung geprüft werden.

Sensorisch-analytische Methoden, die darauf fokussiert sind, sensorische Produktunterschiede objektiv festzustellen, sind diskriminierende Prüfungen. Hierbei müssen in der Regel geringe sensorische Unterschiede zwischen Produkten erkannt und benannt werden.

Daneben gibt es die deskriptiven Prüfungen, die dazu genutzt werden, sensorische Produkteigenschaften (Attribute, Deskriptoren) im Detail zu beschreiben und in ihrer Intensität zu messen. Benötigt werden solche sensorischen Produktprofile (dargestellt häufig in Form von Spiderweb-Diagrammen), um z. B. im Rahmen des Produktmanagements Wettbewerberprodukte mit eigenen Produkten zu vergleichen, eine sensorische Spezifikation für ein neues Produkt und damit auch die Basis (z. B. Referenzmuster, auch „Gold-Standard“ genannt) für die Qualitätskontrolle zu definieren oder sensorische Produktveränderungen bei alternativem Rohstoffeinsatz, bei modifizierten Zutaten, nach längerer Produktlagerung oder aufgrund von Wechselwirkungen mit der Verpackung zu identifizieren.

Sensorische Verfahren, die v. a. auch in der sensorischen Qualitätssicherung eingesetzt werden, um ein gleichbleibendes Qualitätsniveau der produzierten Produkte sicherzustellen, sind z. B. IN/OUT-Tests (Innerhalb/Außerhalb-Prüfung) oder Difference from Control-Tests. Diese werden primär den Unterschiedsprüfungen zugeordnet. Bei beiden Prüfverfahren gibt es jedoch auch deskriptive Varianten. Ziel dieser sensorischen Prüfungen ist es, vorgestellte Produkte gegen eine definierte Referenzprobe (z. B. Produktspezifikation, „Gold-Standard“) zu prüfen und mögliche Abweichungen zu identifizieren. Basierend auf den hierbei erzielten Ergebnissen erfolgt dann eine Produktbewertung auch im Sinne von Produktfreigabe bzw. Nachbearbeitung oder Produktsperre.

Gemäß der Norm DIN 10973 unterscheidet man beim IN/OUT-Test die drei Formen allgemeine [kategorische], skalierte und deskriptive [beschreibende] Prüfung, wobei jeweils geprüft wird, ob die vorgestellte Probe im Vergleich zu einer Referenz innerhalb oder außerhalb der Spezifikation liegt. In der Variante der deskriptiven [beschreibenden] Innerhalb/Außerhalb-Prüfung kann das Verfahren „zusätzlich zur Beurteilung innerhalb/außerhalb um eine vereinfachte Profilierung der Schlüsselattribute erweitert werden“. Der deskriptive Teil folgt dabei zeitlich auf den unterscheidenden Teil der Prüfung. Vergleichbares, die unterscheidenden und beschreibenden Prüfungselemente betreffend, gilt auch für den deskriptiven Difference from Control-Test (DIN 10976).

Während Prüfer und Panelisten für den Einsatz in deskriptiven Panels gezielt im Hinblick auf ihre sensorische Ausdrucks- und Beschreibungsfähigkeit trainiert werden, d. h. die Identifizierung und Benennung von charakteristischen sensorischen Attributen und deren Quantifizierung anhand definierter Skalen verinnerlichen müssen, liegt der Schwerpunkt bei der Qualifizierung diskriminierender Panels auf der selektiven Wahrnehmung von i. d. R. geringen sensorischen Unterschieden. Eine hohe Sensitivität der Sinne und die Kenntnis über relevante Qualitätsparameter spezifischer Produktkategorien sind hier die wesentlichen Qualifizierungsziele. Unabhängig davon, für welchen Bereich in der Lebensmittelsensorik die Prüfer und Panels qualifiziert und eingesetzt werden, sind die regelmäßige Überprüfung und das Monitoring ihrer Leistungsfähigkeit maßgebend für die Qualität und Aussagekraft der zu liefernden sensorischen Ergebnisse.

3. Kriterien zur Messung der Leistungsfähigkeit

Zur Aufrechterhaltung der sensorischen Leistungsfähigkeit sind seitens der Prüfer regelmäßige Teilnahmen an sensorischen Prüfungen sowie ein fortlaufendes Training der Sinne erforderlich. Der Panelleiter muss zu diesem Zweck nicht nur das Training sicherstellen, sondern vielmehr die von ihm in Bezug auf die erforderliche sensorische Prüfkompetenz definierten Kriterien und Leistungsanforderungen kontinuierlich überwachen, um zu gewährleisten, dass die vom Panel gelieferten analytischen Testergebnisse auch valide sind. Bei Abweichungen von der geforderten sensorischen Leistungsfähigkeit muss der Panelleiter mit gezielten Maßnahmen (z. B. fokussiertes Training, sensorische Übungen) entsprechend gegensteuern. Sofern diese Maßnahmen nicht genügen, sind Prüfer von bestimmten sensorischen Prüfungen auszuschließen und/oder Prüferpanels in ihrer Zusammensetzung zu modifizieren.

Bereits in der frühen Sensorikliteratur wird auf die Notwendigkeit zur Überprüfung der Leistungsfähigkeit von Prüferpersonen hingewiesen (siehe z. B. Schweizerisches Lebensmittelbuch seit 1965, Stone und Sidel, 1985). Auch in aktuellen Werken zur Lebensmittelsensorik und Panelqualifizierung wird auf diese Anforderungen verstärkt aufmerksam gemacht (Kemp et al., 2009; Lawless und Heymann, 2010; Raithatha, 2018; Raithatha und Rogers, 2018).

Die verschiedenen Autoren definieren folgende drei Grundanforderungen an die Prüfer- und Panelperformance, an denen sich die Prüferschulung, ebenso wie die Leistungsüberprüfung und das Monitoring der Ergebnisse orientieren müssen. Auch in der DIN EN ISO 8586:2014-05 sind die nachfolgenden Kriterien aufgeführt:

- **Unterscheidungsfähigkeit (Diskriminierfähigkeit):** Die Prüferpersonen sollen unterschiedliche Proben auch als unterschiedlich wahrnehmen (d. h. Unterschiede erkennen, identifizieren bzw. benennen) und skalieren (d. h. die Größe des Unterschieds auf Skalen abbilden, messen) können. Die dahinterstehende Frage lautet: Werden Produktunterschiede immer gleich gut erkannt?
- **Wiederholbarkeit, Vergleichbarkeit (Präzision):** Die Prüferpersonen sollen bei wiederholten Messungen der gleichen Probe unter identischen Bedingungen gleiche oder ähnliche Ergebnisse und Skalenwerte liefern (= Wiederholpräzision: gleiche Probe, gleicher Prüfer, gleicher Ort und gleiche Zeit, d. h. Beurteilung erfolgt in einer Sitzung). Diesem Anforderungsbereich lässt sich auch die Vergleichspräzision zuordnen, bei der zwei oder mehrere Beurteilungen der gleichen Probe zu unterschiedlichen Bedingungen miteinander verglichen werden (= gleiche Probe, gleicher Prüfer, ggf. gleicher Ort aber unterschiedliche Zeit, d. h. mehrere Sitzungen).
Die Frage lautet hier: Sind die Prüfergebnisse innerhalb definierter Spannweiten wiederholbar?
(Achtung: Lebensmittel unterliegen natürlichen Qualitätsschwankungen.)
- **Einheitlichkeit (Homogenität):** Die Prüferpersonen sollen Daten liefern, die mit den Daten der übrigen Panelmitglieder vergleichbar sind oder nahe des Mittelwerts des Panels liegen.
Hierbei fragt man sich: Stimmen die Ergebnisse der Panelmitglieder weitgehend (innerhalb einer definierten Spannweite) überein oder gibt es gravierende Ausreißer?
(Anmerkung: Eine Null-Streuung ist in der Praxis nicht möglich.)

Generell gelten diese Kriterien zum Monitoring der sensorischen Leistungsfähigkeit für alle Prüferpersonen, die an objektiven sensorisch-analytischen Untersuchungen teilnehmen, sowie für das gesamte Panel. Unabhängig davon können die Kriterien und Faktoren auch während der Ausbildungsphase des Panels verwendet werden, um den aktuellen Schulungsstand zu dokumentieren.

In der Überwachungspraxis und im Monitoring allerdings variieren die hinter diesen Kriterien stehenden Parameter und ihre jeweiligen Grenzwerte bzw. die Gewichtung in Abhängigkeit vom jeweiligen Einsatzzweck der Prüferpersonen bzw. Panels. In der Produktentwicklung und -optimierung liegt der Fokus eher im deskriptiven Bereich, sodass Prüfer und Panels nicht nur die Fähigkeit besitzen müssen, Unterschiede und Intensitäten zu messen, sondern v. a. auch wortgewandt

und sicher in der Anwendung der produktspezifischen sensorischen Sprache (Terminologie) sein müssen. Für die Qualitätssicherung (QS) sind methodisch eher diskriminierende Ansätze gefragt, d. h. Prüfer und Panels benötigen eine hohe Sensitivität und Unterscheidungsfähigkeit. Reine QS-Panels, die v. a. für IN/OUT-Tests eingesetzt werden, wiederum haben nochmals weiterführende Anforderungen u. a. auch an die lebensmitteltechnologische und warenkundliche Kompetenz, wenn es darum geht v. a. Abweichungen von Referenzproben zu identifizieren.

4. Praktische Umsetzung der Leistungsüberprüfung sensorischer Prüfpersonen und Panels

Zum Einstieg in die praktische Umsetzung einer Leistungsüberprüfung sind verschiedene Aspekte zu berücksichtigen. Sicherlich gelten in vielen Betrieben betriebsspezifische Anforderungen u. a. seitens des Betriebsrates den Datenschutz oder seitens der IT-Abteilung den Einsatz von Software-Tools betreffend, die zu Einschränkungen bei der Datenauswertung führen können. Sofern also aus Datenschutzgründen keine individuellen Prüfergebnisse der Prüfer, auch nicht in anonymisierter Form, zu Zwecken der Leistungsauswertung und zu deren Monitoring aber auch zur Prüfermotivation (d. h. zur Abbildung der eigenen sensorischen Kompetenzen) genutzt werden dürfen und auch der Abschluss von Vertraulichkeitsvereinbarungen o. ä. Formen einer „gesicherten Datennutzung“ nicht umsetzbar ist, ermöglichen – wenngleich mit limitierter Aussagekraft – komprimierte Daten in Form der Auswertung von Gruppenergebnissen Anhaltspunkte für die Identifikation von leistungsbezogenen Schwachstellen und Verbesserungspotenzialen. Aufbauend auf diesen lassen sich gleichfalls Schulungsmaßnahmen planen und Trainings realisieren, um die sensorische Leistungsfähigkeit der Prüfergruppen weiter zu verbessern und zu optimieren. Welche Möglichkeiten diesbezüglich angemessen und gestattet sind, muss der jeweilige Pannelleiter anhand der betriebsspezifischen Gegebenheiten eruiieren sowie letztlich dann individuell bewerten und entscheiden.

Aus allgemeiner fachlicher Sicht lassen sich hinsichtlich der praktischen Umsetzung einer Leistungsüberprüfung und -bewertung von Prüfpersonen bzw. Panels generell verschiedene Betrachtungsweisen unterscheiden (Rogers, 2018), die bei der jeweiligen Planung im Vorfeld zu berücksichtigen sind:

1. Untersuchungsgegenstand:

Es gilt festzulegen, ob das Monitoring einer einzelnen Prüferleistung betrachtet werden soll oder vielmehr die Überprüfung einer „gebündelten“ Panelleistung zu analysieren ist. Diese Einzelergebnisse können auch untereinander in Beziehung gesetzt werden. D. h. Vergleiche von Prüfer und Panel oder Panel mit Panels aus anderen Unternehmen oder Unternehmensstandorten.

2. Verfahrensweise:

a) Laufende/neue Projekte:

Es können Prüfer- und Paneldaten aus laufenden Projekten für weiterführende Auswertungen die Leistungsfähigkeit betreffend genutzt oder Ergebnisse aus speziell für das Monitoring neu entwickelten Projekten zur Beurteilung herangezogen werden.

b) Vorhandene/modifizierte Produktproben:

Die zu analysierenden Produktproben können entweder Marktmuster oder Produkte aus der eigenen Produktion sein. Darüber hinaus können Produkte zu Übungszwecken auch gezielt modifiziert („gespiket“) werden, z. B. durch Verstärkung des Vanilleeindrucks bei Desserts durch einen höheren mengenmäßigen Zusatz von Vanillearoma o. Ä. Dies hat den Vorteil der detaillierten Produktkenntnis und stellt eindeutig sicher, dass es Unterschiede bei bestimmten Merkmalseigenschaften gibt. Diese werden dann als Schlüsselmessgrößen für die Einordnung der Leistungsfähigkeit verwendet, die seitens des Prüfers bzw. Panels erreicht werden müssen.

3. Zeitverlauf:

Eine Leistungsüberprüfung von einzelnen Prüfpersonen oder Panels kann, wie in Tabelle 1 dargestellt, punktuell (am Ende einer Schulungseinheit) oder im Rahmen eines Monitorings auch über einen längeren zeitlichen Verlauf (Zeitspanne) betrachtet werden.

	Leistungsbewertung zum definierten Zeitpunkt	Leistungsüberwachung als Trendbewertungen über einen längeren Zeitraum
Leistungsbewertung im Rahmen der regelmäßigen Prüfer- und Paneltätigkeit	Die Daten zur punktuellen Leistungsbewertung bei Prüfern und Panels können aus laufenden Projekten entnommen und gezielt ausgewertet werden.	Die Daten zur fortlaufenden Leistungsüberwachung bei Prüfern und Panels können aus bereits länger laufenden Projekten entnommen und gezielt ausgewertet werden.
Leistungsbewertung im Rahmen von speziell definierten Performance-Tests	Definierte Anforderungen/Leistungskriterien werden mittels speziell definierter Projekte punktuell überprüft.	Definierte Anforderungen/Leistungskriterien werden mittels speziell definierter Projekte in einem fortlaufenden Prozess überprüft.

Tabelle 1: Ansätze zur Umsetzung einer Leistungsbewertung (in Anlehnung an Rogers, 2018)

Die sensorische Leistungsfähigkeit muss laufend trainiert werden. Idealerweise sind Prüfpersonen regelmäßig und mit nur kleinen Unterbrechungen im Einsatz. In einem solchen Fall können die Ergebnisse aus laufenden Projekten für die Leistungsüberprüfung herangezogen werden. Es empfiehlt sich dann auch diesbezüglich ein entsprechendes Konzept und eine passende Dokumentation zu erarbeiten, die die getrennte Datenauswertung vom eigentlichen sensorisch-analytischen Projekt aufzeigt. Neben Auswertungen aus laufenden Projekten ist es empfehlenswert, mit zusätzlichen Prüfungen ggf. im Rahmen von regelmäßigen Schulungen oder auch mit kurzen Tests vor den eigentlichen Projektsitzungen die Leistungsfähigkeit zu ermitteln.

Während die Überprüfung der Leistungsfähigkeit im Bereich der deskriptiven Prüfungen auch über die DIN EN ISO 11132:2017-10 weitgehend beschrieben und standardisiert ist, so dass man für die praktische Umsetzung darauf aufbauend unternehmensspezifische und an die jeweilige Produktkategorie adaptierte Möglichkeiten ableiten kann, obliegt die Konzeption von Maßnahmenpaketen im Bereich der „Diskriminierenden Prüfungen“ dem zuständigen Panelleiter, der sich ggf. entsprechende Informationen aus verschiedenen Literaturquellen zusammenstellen muss. Unabhängig davon, ob die Prüfer im Rahmen von Unterschiedsprüfungen oder bei deskriptiven Prüfungen eingesetzt werden, gelten – wie zuvor beschrieben – zur Überprüfung der Leistungsfähigkeit stets die drei Kriterien Unterscheidungsfähigkeit, Wiederhol-/Vergleichbarkeit und Homogenität als Basis.

Sofern die **Unterscheidungsfähigkeit/Diskriminierfähigkeit** überprüft werden soll, sind den Prüfern mehrere ähnliche Proben zum Test zu reichen. Dabei ist bei der Auswahl von Proben unbedingt zu beachten, dass die Unterschiede nicht zu klein und auch nicht zu groß sind, denn ansonsten besteht die Gefahr, dass man im Ergebnis eine (zu) gute oder (zu) schlechte Diskriminierfähigkeit interpretiert. Ein Prüfer oder das Panel diskriminiert schlecht, wenn a) nie ein sensorischer Unterschied gefunden wird oder b) die eigentlich unterschiedlichen Produkte im Hinblick auf die sensorischen Attribute und/oder Intensitäten als sehr ähnlich bewertet werden. Zur Verbesserung der Diskriminierfähigkeit tragen Schulungen bei, die sowohl die Sensitivität der Sinne (u. a. Rangordnungsprüfungen, Dreiecksprüfungen) trainieren als auch die Skalennutzung (d. h. die Abbildung der Größe des Unterschieds).

Die Qualität der **Wiederholbarkeit** wird durch den Vergleich der Prüfergebnisse sowohl einzelner Prüfer mit sich selbst als auch des gesamten Panels mit sich selbst deutlich. Hierbei sollten identische Proben auch bei einer erneuten Prüfung identisch bewertet werden, d. h. gleiche oder ähnliche Ergebnisse und Skalenwerte liefern. Um diese Fähigkeit zu überprüfen, eignen sich z. B. „**verdeckte Doppelproben**“, die in eine Sitzung eingebaut und den Prüfern als gleichwertiges Produkt (gleiche Charge) gereicht werden (d. h. gleiche Prüfer/Panels, gleiches Produkt, gleicher Ort, gleiche Zeit). Eine weitere Möglichkeit diesbezüglich stellen „**Wiederholungs-Sessions**“ dar. Dabei sind seitens der Prüfer mehrere gleichartige Sitzungen zu absolvieren, d. h. die definierten sensorischen Tests mit den gleichen Proben (gleiche Charge, sonst vergleichbare Proben) werden in bestimmten Abständen über einen längeren Zeitraum erneut durchgeführt (d. h. gleiche Prüfer/Panels, gleiches Produkt, gleicher Ort, unterschiedliche Zeit).

Eine gute **Homogenität** ist dann gegeben, wenn der Grad der Übereinstimmung sensorischer Beurteilungen von Prüfern im Vergleich zu anderen Prüfern sowie dem gesamten Panelergebnis hoch ist. Dies lässt sich gut über eine Datenauswer-

tung erfassen. In diesem Kontext gibt es auch die Möglichkeit des Datenvergleichs mit anderen Panels. Solche **Ringversuche** (Proficiency Tests) werden durchgeführt, um das jeweils angewendete Untersuchungsverfahren zu überprüfen und die Ergebnisse des eigenen Panels mit den Ergebnissen anderer Panels zu vergleichen. Ringversuche, die in der chemischen Analytik bereits seit längerer Zeit im Rahmen der Guten Laborpraxis (GLP) etabliert sind, können in der Lebensmittel-sensorik produktspezifisch und unternehmensintern (z. B. Vergleich von Panels verschiedener Unternehmensstandorte) oder branchenspezifisch (z. B. überbetrieblicher Vergleich von Panels) organisiert werden. Von der organisierenden Stelle werden dazu für alle Prüfer einheitliche Proben verschickt, die von den teilnehmenden Panels innerhalb eines gegebenen Zeitraumes nach den gleichen, zuvor definierten Prüfmethoden zu untersuchen sind (d. h. unterschiedliche Prüfer/Panels, gleiche Produkte, unterschiedliche Orte, unterschiedliche Zeiten). Die eingereichten Ergebnisse können nach den bereits beschriebenen Kriterien zur Messung der Leistungsfähigkeit eines sensorischen Panels ausgewertet werden.

5. Fallbeispiele zur Beurteilung der Prüfer- und Panelperformance

Der Einsatz speziell entwickelter professioneller Softwareprogramme in der sensorischen Analyse hat den Vorteil, dass sämtliche Daten für die Auswertung in digitaler Form vorliegen und somit unmittelbar weiterverarbeitet werden können. Softwareprogramme wie z. B. FIZZ oder Compusense bieten dabei die Möglichkeit, den Prüfprozess, je nach Methodik, digital zu hinterlegen und helfen gleichzeitig Eingabe- und Übernahmefehler zu vermeiden, in dem z. B. die Eingabe von Werten erzwungen wird (forced choice) und ein Weiterarbeiten erst nach erfolgter Antwort möglich ist. So gesammelte digitale Daten sind einerseits die Grundlage für die statistische Auswertung sensorischer Prüfverfahren und andererseits auch die Basis für die statistischen Auswertungen zur Beurteilung der Prüfer- und Panelperformance. Für die Auswertung sensorischer Daten existieren viele verschiedene statistische Methoden. Keines dieser Verfahren und keine Art der grafischen Darstellung ist alleine in der Lage, sämtliche in „Sensorik-Daten“ vorhandenen Informationen umfassend aufzuarbeiten. Man benötigt stets eine Kombination verschiedener Sichtweisen und Perspektiven, welche durch verschiedene statistische Auswertungen und grafische Darstellungen generiert werden sollten.

Die nachfolgenden Fallbeispiele aus den Einsatzbereichen „Diskriminierende Prüfungen“, „Diskriminierende und Deskriptive Prüfungen i. S. v. Deskriptiven Difference from Control- bzw. Deskriptiven IN/OUT-Tests“ und „Deskriptive Prüfungen“ möchten, unter exemplarischer Einbindung verschiedener Auswertungstools, einen Einblick in die sich bietenden Möglichkeiten zum Monitoring der jeweiligen Prüfer- und Panelperformance geben. Seitens des Unternehmens bzw. des verantwortlichen Pannelleiters ist stets betriebspezifisch zu prüfen, welche Möglichkeiten in Anspruch genommen werden sollten. Sofern unternehmensintern der Einsatz entsprechender am Markt angebotener Sensorik-Softwareprogramme nicht oder noch nicht möglich ist, bieten sich für Pannelleiter auch einfache Auswertungsmöglichkeiten unter Einsatz von MS-Excel in der Standardversion oder über XLSTAT, ein vertiefendes und kostenpflichtiges Add-in Modul in MS-Excel.

5.1 Fallbeispiel 1: „Diskriminierende Prüfungen“ am Beispiel Dreieckstest

Besonders wichtig für Panels, die insbesondere Unterschiedsprüfungen anwenden, ist eine grundsätzlich hohe Sensitivität der einzelnen Prüfpersonen sowie deren Fähigkeit, Unterschiede zwischen Prüfmustern identifizieren zu können. Bei hoher Diskriminierfähigkeit einzelner Prüfpersonen resultiert eine hohe Diskriminierfähigkeit des gesamten Panels.

Zur besseren Einschätzung der Leistungsfähigkeit von Prüfern und Panels bietet es sich an, diskriminierende Prüfungen, z. B. Dreiecksprüfungen, Paarweise Vergleichsprüfungen, Duo/Trio-Prüfungen etc., mit unterschiedlichen Schwierigkeitsgraden durchzuführen. Bei Betrachtung der Einzelergebnisse erkennt man die Fähigkeit der Prüfpersonen Unterschiede korrekt zu identifizieren. Datensätze gesamter Panels sind i. d. R. binomial verteilt und können mit Hilfe von in den jeweiligen Normen abgebildeten Tabellen ausgewertet werden oder aber mittels Einsatz statistischer Programme, wie beispielsweise der Add-in Software XLSTAT für MS-Excel.

Summary of selected options:	
Test	Triangle Test
Number of assessors	15
Prop. of correct answers	0.533
Guessing probability	0.333
Sensory discrimination tests:	
Prop. of discrim.	0.300
z	1.590
p-value	0.056
alpha	0.050
Power	0.401
Test interpretation:	
H0: The two products are similar.	
Ha: The two products are different.	
As the computed p-value is greater than the significance level $\alpha = 0.05$, one cannot reject the null hypothesis H0.	
The risk to reject the null hypothesis H0 while it is true is 5.60 %.	

Tabelle 2: Beispiel Auswertung Dreiecksprüfung mit XLSTAT (Guessing Model), 15 Prüfer, 8 richtige Antworten

Summary of selected options:	
Test	Triangle Test
Number of assessors	21
Prop. of correct answers	0.524
Guessing probability	0.333
Sensory discrimination tests:	
Prop. of discrim.	0.286
z	1.793
p-value	0.037
alpha	0.050
Power	0.415
Test interpretation:	
H0: The two products are similar.	
Ha: The two products are different.	
As the computed p-value is lower than the significance level $\alpha = 0.05$, one should reject the null hypothesis H0, and accept the alternative hypothesis Ha.	
The risk to reject the null hypothesis H0 while it is true is lower than 3.65 %.	

Tabelle 3: Beispiel Auswertung einer Dreiecksprüfung mit XLSTAT (Guessing Model), 21 Prüfer, 11 richtige Antworten

Nachfolgend werden zwei Beispiele für die Auswertung diskriminierender Datensätze und die Interpretation resultierender Ergebnisse dargestellt. Konkret handelt es sich um Ergebnisse einer **Dreiecksprüfung (DIN EN ISO 4120)**. Unter Annahme einer Irrtumswahrscheinlichkeit von $\alpha = 0.05$ wird für Prüfer und Panel die Anzahl richtig gelöster Tests pro Test in Prozent berechnet.

In Tabelle 2 ist zu erkennen, dass der p-Wert größer als 0.05 (5 % Irrtumswahrscheinlichkeit) ist und bei 0.056 liegt. D. h. der Unterschied zwischen den Prüfproben wurde nicht signifikant erkannt. Das Risiko einen Unterschied zu finden, obwohl keiner da ist, liegt in diesem Fall bei 5.6 % und der Anteil Unterscheider p_d (proportion of discriminators) liegt bei 30 %.

In Tabelle 3 lässt sich ablesen, dass der p-Wert kleiner als 0.05 (5 % Irrtumswahrscheinlichkeit) ist und bei 0.037 liegt. D. h. der Unterschied zwischen den Prüfproben wurde signifikant erkannt. Das Risiko einen Unterschied zu finden, obwohl keiner da ist, liegt in diesem Fall bei 3.7 % und der Anteil Unterscheider p_d (proportion of discriminators) liegt bei 28.6 %.

Führt man dann z. B. Dreiecksprüfungen in mehreren Sitzungen über einen längeren Zeitraum in gleicher Form durch, so könnte man das in Abbildung 2 angeführte Monitoring-Diagramm erstellen, welches die fachliche Fortentwicklung der Unterscheidungsfähigkeit eines Panels anhand des %-Anteils richtiger Antworten aufzeigt. In diesem Beispiel haben $n = 21$ Prüfer in fünf gleichartigen Sitzungen über fünf Wochen einen Dreieckstest absolviert. Im Durchschnitt aller Prüfungen waren 69 % der Antworten richtig (rote Linie). Durch systematisches Training konnte damit die Leistungsfähigkeit bei der Diskriminierung seitens der Prüfer sukzessive gesteigert werden.

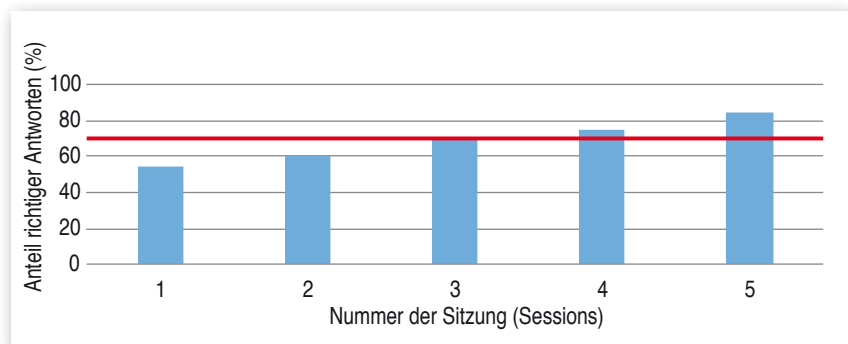


Abbildung 2: Monitoring der Diskriminierfähigkeit eines Panels

5.2 Fallbeispiel 2: „Diskriminierende Prüfungen und Deskriptive Prüfungen“ am Beispiel des „Deskriptiven Difference from Control-Tests“

Im Rahmen eines Deskriptiven Difference from Control-Tests von Getränken (vgl. DLG-Expertenwissen 2-2013 und 2-2013 sowie DIN 10976:2016-08) bei dem ein aus vier Prüfern bestehendes Panel anhand einer 6-Punkte Unterschiedsskala (Bewertungen von 1 = keine Abweichung, vollständige Übereinstimmung bis 6 = sehr große Abweichung, nicht tolerierbar) jeweils zehn vorgestellte Produktproben analysierte, wurden die in den Abbildungen 3 und 4 dargestellten Ergebnisse erzielt. Während im Beispiel 1 (Abbildung 3) die vier geschulten und routinierten Prüfer relativ homogen beurteilten, d. h. die Abweichungen der Ergebnisse des Prüfers 1 vom Gruppenergebnis um die $\pm 0,5$ Punkte liegen, stellt sich die Situation im Beispiel 2 (Abbildung 4) anders dar. Hier weichen die Gruppenergebnisse zum Teil erheblich von den Prüfergebnissen des Prüfers 2 ab. Letzterer ist ein neuer Prüfer und ggf. noch nicht sehr versiert entweder in der fachlichen Beurteilung der Produkte oder in der Anwendung der Skala. Hier besteht Schulungsbedarf entweder im Bereich der Skalennutzung oder hinsichtlich der Ermittlung sensorischer Unterschiede.

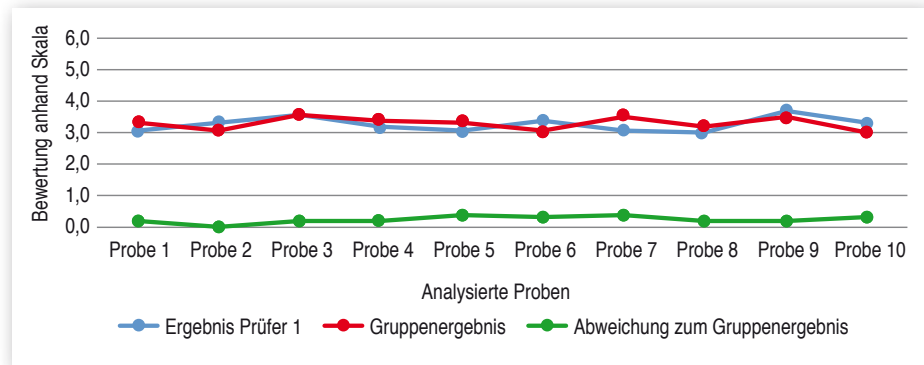


Abbildung 3: Beispiel 1: Prüfermonitoring „Difference from Control-Test“ mit MS-Excel; $n = 4$ routinierte Prüfer, bewertet wurden 10 Getränkeproben

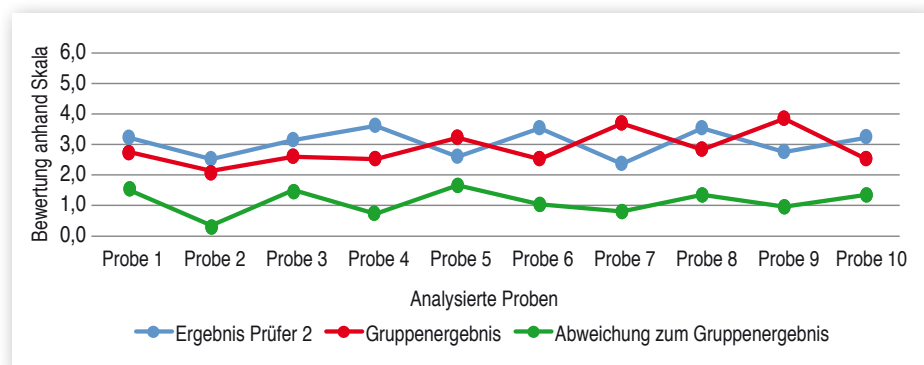


Abbildung 4: Beispiel 2: Prüfermonitoring „Difference from Control-Test“ mit MS-Excel; $n = 4$ Prüfer, darunter 3 routinierte Prüfer und 1 neuer Prüfer, bewertet wurden 10 Getränkeproben

5.3 Fallbeispiel 3: „Deskriptive Prüfung“

In der DIN EN ISO 11132:2017:10 sind Ausführungen hinsichtlich des Monitorings von Prüfern und Panels, die bei deskriptiven Prüfungen eingesetzt werden, publiziert. Die hier beschriebene Vorgehensweise erfordert detaillierte statistische Kenntnisse und den Einsatz einer Software, die die Datenverarbeitung und Ergebnisermittlung erleichtert bzw. automatisiert. Die unter 5.3.2 genannten Beispiele aus der Software PanelCheck sind Bestandteile der Ausführungen in dieser Norm. Generell sollten sich Auswertungen von Pannelleistungen an diesen Vorgaben orientieren. Für Einsteiger in diese Thematik oder für Panelleiter, die einen schnellen Überblick über die Leistungen ihres deskriptiven Panels bzw. die einzelner Prüfer erhalten möchten, eignen sich auch die nachfolgend dargestellten Beispiele unter Einsatz von MS-Excel, wobei die Kriterien zur Prüfer- und Panelperformance aus der DIN EN ISO 8586:2014-05, d. h. die Unterscheidungsfähigkeit (Diskriminierfähigkeit), Wiederhol- bzw. Vergleichbarkeit (Präzision) und Einheitlichkeit (Homogenität) als Basis gelten.

5.3.1 Auswertungen und Darstellungen unter Einsatz von MS-Excel

5.3.1.1 Diskriminierfähigkeit (Unterscheidungsfähigkeit)

Um zu prüfen, ob die Prüfpersonen unterschiedliche Proben auch als unterschiedlich wahrnehmen (d.h. Unterschiede erkennen, identifizieren bzw. benennen) und skalieren (d.h. die Größe des Unterschieds auf Skalen abbilden, messen) können, eignet sich der Einsatz von bekannten Referenzproben. Nach der in der DIN EN ISO 13299:2016-09 beschriebenen Vorgehensweise zur Erstellung eines „Profils der Abweichung vom Referenzstandard“ werden dabei die zu bewertenden Proben paarweise gereicht. Die eine wird als Referenz, die andere als Probe 1 bezeichnet. In diesem Beispiel wurde die Probe 1 mit einem Zusatz von Salz im Geschmack modifiziert. Dem Prüfungsleiter sind die Produktprofile beider Proben bekannt, so dass er die Bewertungen des Prüfers und das letztliche Prüfergebnis einschätzen kann. Der Prüfer ist nun aufgefordert, nach einer ihm bekannten, vorgegebenen Deskriptorenliste und auf Basis einer ihm bekannten Skala die Probe 1 sowohl qualitativ als auch quantitativ im Vergleich zur Referenzprobe zu beschreiben. Die dahinterstehende Frage lautet: Werden Produktunterschiede erkannt und treffend beschrieben bzw. bei Wiederholungsprüfungen immer gleich gut erkannt?

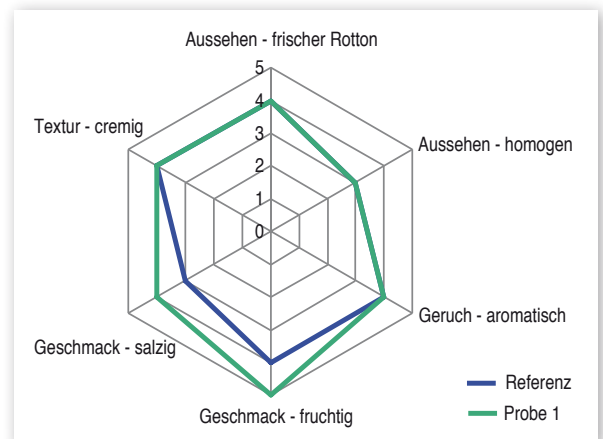


Abbildung 5: Prüfermonitoring einer Profilanalyse eines Tomatensaftes, Gute Diskriminierfähigkeit des Prüfers Nr. 5 (Sitzung mit Referenzprobe)

Auswertung mit MS-Excel; $n = 1$ Prüfer, bewertet wurden 2 Proben, darunter 1 Referenz

In Abbildung 5 ist ein Beispiel für eine gute Diskriminierfähigkeit des Prüfers dargestellt. Der Prüfer Nr. 5 hat alle Deskriptoren und ihre Bewertungen gut erkannt und den Salzzusatz bei Probe 1 beschrieben. Dass dieser sich auch ggf. im Geschmack auf die Fruchtigkeit auswirkt, hat er anhand der Intensität dokumentiert. Um die Wiederholbarkeit der Ergebnisse zu überprüfen, müsste man wie unter 5.3.1.2 beschrieben, mehrere gleichartige Sitzungen durchführen und die Ergebnisse entsprechend analysieren.

5.3.1.2 Wiederhol- bzw. Vergleichbarkeit (Präzision) des Panels und der Prüfer

Um die **Wiederholbarkeit der Leistungen eines Panels als Gruppe** zu überprüfen bestehen, wie in den vorangegangenen Kapiteln dargelegt, verschiedene Möglichkeiten über den Einsatz von Doppelproben oder über die Durchführung von Wiederholungsprüfungen. In diesem Fallbeispiel prüfen acht Panelisten zwei Tomatensäfte nach der in der DIN EN ISO 13299:2016-09 beschriebenen Vorgehensweise zur Erstellung eines „Quantitativ beschreibenden Profils“. Sie haben dazu im Vorfeld die wesentlichen Deskriptoren festgelegt sowie sich für den Einsatz einer Skala von 0 – 5 entschieden. Dabei bedeutet 0, dass die genannte Eigenschaft „nicht erkennbar“ bzw. 5, dass sie „sehr stark erkennbar“ ist. Das Panel befindet sich im Aufbau, so dass erfahrene und weniger erfahrene Prüfer dabei sind. Im nachfolgenden Beispiel wurde im Anschluss an die erste Sitzung nach zwei Wochen eine Wiederholungssitzung durchgeführt (Vergleichspräzision), d.h. gleiche Produktcharge, gleiche Zeit, gleiche Prüfer, gleicher Ort aber andere Zeit (zwei Wochen

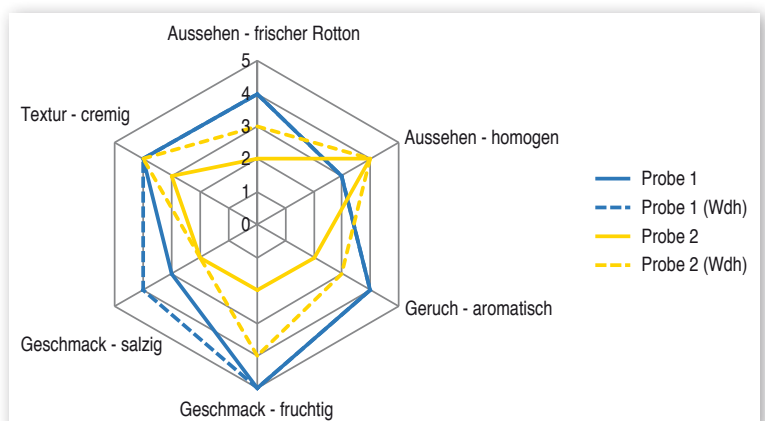


Abbildung 6: Panelmonitoring einer Profilanalyse eines Tomatensaftes: Wiederholbarkeit des Panels als Gruppe (Sitzung 1) Auswertung mit MS-Excel; $n = 8$ Prüfer, darunter 6 routinierte Prüfer und zwei neue Prüfer, bewertet wurden zwei verschiedene Proben zu unterschiedlichen Zeiten

später). Aus den einzelnen Prüfergebnissen wurde jeweils der Panelmittelwert errechnet und grafisch dargestellt. Je besser die beiden Bewertungen (Erstbewertung und Wiederholungsbewertung) übereinander liegen, desto besser wiederholt das Panel. Solche Abbildungen können dem Panel gut als Rückmeldung gezeigt werden.

Die Ergebnisse beider Sitzungen sind in Abbildung 6 visualisiert, wobei die durchgezogenen Linien die Panelmittelwerte der ersten Sitzung und die gestrichelten Linien die Panelergebnisse aus der Wiederholungssitzung aufzeigen. Hier lässt sich sehr gut erkennen, dass die Gruppenergebnisse unterschiedlich ausfallen. Während das Messergebnis der Prüfer bei Probe 1 (blaue Linien) eine gute Wiederholbarkeit zeigt (Spiderweb-Linien sind fast überall deckungsgleich), liegen die Panelbewertungen bei der Probe 2 (gelben Linien) stärker auseinander. Hier ist die Wiederholbarkeit während der Wiederholungssitzung weniger gut, die beiden Spiderweb-Linien sind kaum deckungsgleich. Im Zuge der Ursachenforschung gilt es sowohl die Sinnesphysiologie als auch die Skalennutzung der Prüfer detaillierter zu betrachten und entsprechende Schulungsmaßnahmen zu konzipieren.

Um die Leistungsfähigkeit im Zuge der **Wiederholbarkeit von Prüfern** zu beurteilen, bietet es sich an, in einer Sitzung Doppelproben einzubinden. Die nachfolgenden Beispiele in den Abbildungen 7 und 8 sind diesem Fall gewidmet. Hierbei wurden dem Prüfer Nr. 3 insgesamt fünf Proben von Tomatensäften gereicht, vier davon unterschieden sich sensorisch, eine Probe war als Doppelprobe zweimal enthalten.

In Abbildung 7 ist der Fall einer schlechten Wiederholbarkeit des Prüfers Nr. 3 dargestellt. Dieser Prüfer hat die Probe 1 (blaue Linie) und die Probe 5 (hellblaue Linie) sehr unterschiedlich bewertet.

Dahingehend stellt die Abbildung 8 den gegenteiligen Fall, nämlich eine gute Wiederholbarkeit eines Prüfers dar. Der Prüfer Nr. 4 hat die Probe 1 (blaue Linie) und die Probe 5 (hellblaue Linie) nahezu identisch bewertet.

5.3.1.3 Homogenität (Einheitlichkeit des Panels)

Bezüglich der Homogenität ist zu überprüfen, ob die Bewertungen der jeweiligen Prüfer mit den übrigen Panelmitgliedern vergleichbar sind und/oder nahe des Mittelwerts des Panels liegen bzw. ob es ggf. Ausreißer gibt. In diesem Fallbeispiel prüfen acht Panelisten einen Tomatensaft nach der in der DIN EN ISO 13299:2016-09 beschriebenen Vorgehensweise zur Erstellung eines „Quantitativ beschreibenden Profils“.

Sie haben dazu im Vorfeld die wesentlichen Deskriptoren festgelegt sowie sich für den Einsatz einer Skala von 0 – 5 entschieden. Dabei bedeutet 0, dass die genannte Eigenschaft „nicht erkennbar“ bzw. 5, dass sie „sehr stark erkennbar“ ist. Sechs der acht Panelisten sind routinierte Prüfer, zwei davon sind noch relativ neue Prüfer.

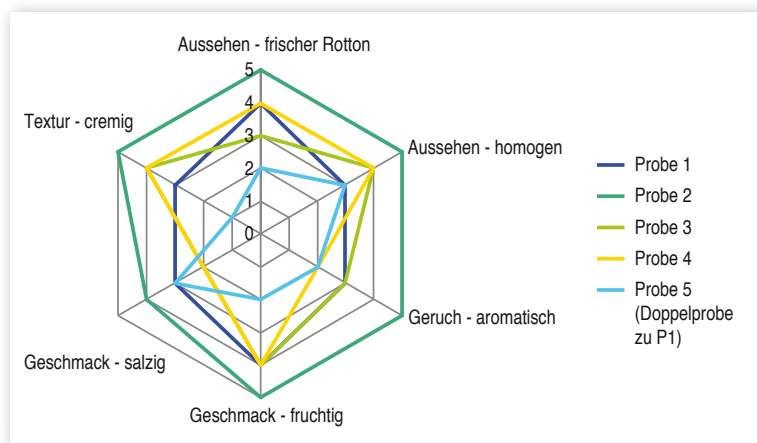


Abbildung 7: Prüfermonitoring einer Profilanalyse eines Tomatensaftes: Schlechte Wiederholbarkeit des Prüfers Nr. 3 (Sitzung mit Doppelprobe) Auswertung mit MS-Excel; n = 1 Prüfer, bewertet wurden 5 Proben, darunter 1 Doppelprobe

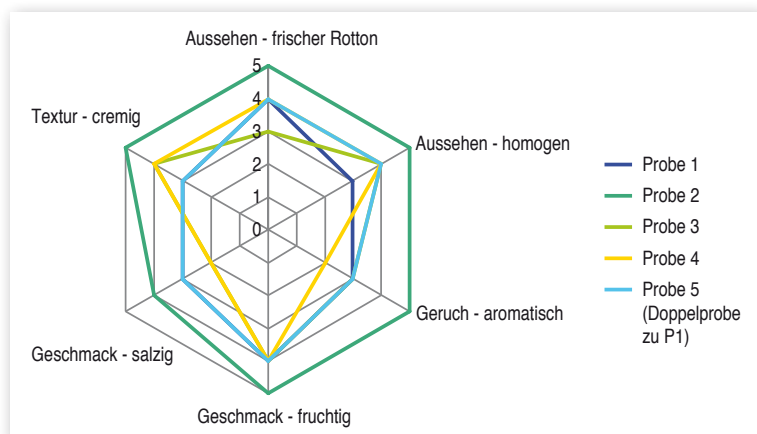


Abbildung 8: Prüfermonitoring einer Profilanalyse eines Tomatensaftes: Gute Wiederholbarkeit des Prüfers Nr. 4 (Sitzung mit Doppelprobe) Auswertung mit MS-Excel; n = 1 Prüfer, bewertet wurden 5 Proben, darunter 1 Doppelprobe

In Abbildung 9 sind in einem Spiderweb die Ergebnisse der einzelnen Prüfer sowie das Panelergebnis im Durchschnitt (rote Linie) dargestellt. Wie daraus zu entnehmen ist, fallen insbesondere der Prüfer 2 und der Prüfer 4 dahingehend auf, dass ihre Prüfergebnisse von denen der anderen Prüfer und vom Paneldurchschnitt abweichen. Während der Prüfer 2 (violette Linie) die Produkteigenschaften weitgehend als „stark erkennbar“ bis „sehr stark erkennbar“, also als sehr intensiv bewertet (Skalenbereich 4-5), findet der Prüfer 4 (gelbe Linie) die Tomatensaftprobe als wenig bis schwach intensiv, denn seine Angaben bewegen sich im Skalenbereich zwischen 0–2. Das Panel arbeitet damit nicht

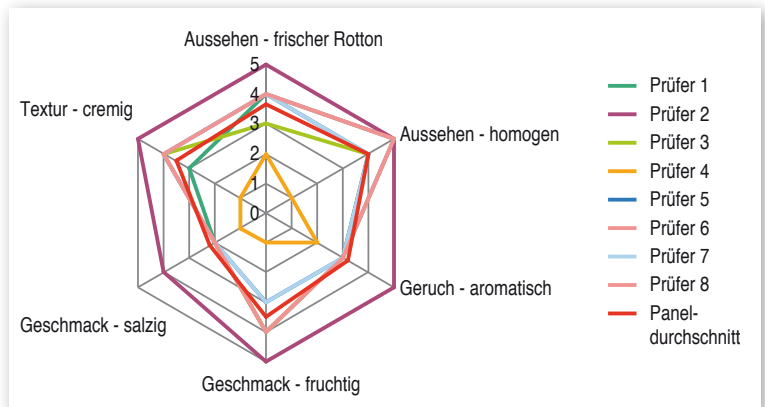


Abbildung 9: Panelmonitoring einer Profilanalyse eines Tomatensaftes: Homogenität des Panels

Auswertung mit MS-Excel; $n = 8$ Prüfer, darunter 6 routinierte Prüfer und 2 neue Prüfer, bewertet wurde 1 Probe

homogen und es bedarf weiterer Schulungen. Die unterschiedlichen Bewertungen durch die Prüfer 2 und 4 können durch Defizite in der Sinnesphysiologie bedingt sein oder aber auch dadurch, dass die Prüfer noch nicht so sicher im Umgang mit der gewählten Skala sind. Durch ein entsprechendes Screening und Training der sensorischen Leistungsfähigkeit und über ein anschließendes intensives Skalentraining kann die einzelne Prüferleistung verbessert und der Panelleistung angeglichen werden. Damit lässt sich die Homogenität verbessern.

Die zuvor mittels MS-Excel dargestellten Fallbeispiele in Form von Spinnennetzen zeigen einfache Auswertungsmöglichkeiten auf, auch indem sie den Blick auf wenige Produkte bzw. eine überschaubare Anzahl von Deskriptoren richten. In der Praxis stellen sich Projekte jedoch häufig komplexer dar, so dass es vor allem beim Vergleich mehrerer Produkte und dem Vorliegen vieler Attribute und größerer Prüfergruppen schwieriger wird, die Daten über die o.g. Darstellungsform zu visualisieren und Prüfmuster zu identifizieren. Hierbei kann dann die mehrdimensionale Darstellung im Koordinatensystem (vgl. PanelCheck) zielführender sein.

5.3.2 Auswertungen und Darstellungen mittels „PanelCheck“

Eine mögliche Lösung ist das kostenlose (Freeware) **Software-Paket „PanelCheck“** (www.panelcheck.com), welches auf dem „open-source“ Statistikprogramm „R“ basiert und auf deskriptive Prüfungen zugeschnitten ist. Hier werden, nachdem die eigenen Prüfdaten in einer qualitativ den Programmanforderungen entsprechenden elektronischen Form (z. B. Excel) überspielt wurden, über automatisierte Prozesse und auf Basis verschiedener statistischer Methoden entsprechende Auswertungen generiert. Letztere zeigen sowohl als Zahlenwerte aber auch in Form von grafischen Darstellungen ein Gesamtbild der Leistungsfähigkeit/Performance eines Panels resp. einzelner Prüfer (Panelisten) und bieten für (fortgeschrittene) Panelleiter eine ideale Unterstützung im Panelmanagement-Prozess. Das Programm ist anwenderfreundlich und bedingt durch die voreingestellte automatisierte Ermittlung von statistischen Kenngrößen auch für „Nichtstatistiker“ nutzbar. Die Publikationen von Tomic et al. (2007, 2010) geben einen tieferen Einblick in die sich bietenden Auswertungen und statistischen Kenngrößen, die auf Empfehlungen in aktuellen DIN EN ISO-Normen basieren.

Die nachfolgenden Beispiele zeigen eine Auswahl von Auswertungsmöglichkeiten die sensorische Leistungsfähigkeit deskriptiver Prüfer und Panels betreffend, die sich mittels „PanelCheck“ generieren lassen und v. a. in der DIN EN ISO 11132:2017:10 empfohlen werden. Die hier dargestellten Abbildungen bilden verschiedene Sitzungen und damit unterschiedliche Fallbeispiele ab.

5.3.2.1 Diskriminierfähigkeit

Die **Diskriminierfähigkeit** eines Prüfers bzw. des Panels lässt sich mit den **F plots** in der PanelCheck-Software überprüfen, die nach dem Rohdatentransfer seitens des Panelleiters entsprechend ausgewählt werden können (Abbildung 10). Je größer der F-Wert bzw. je weiter der p-Wert unter dem vorgegebenen Signifikanzniveau liegt, desto höher ist die Fähigkeit einer Prüfperson zwischen den Produkten zu unterscheiden. Weist das gesamte Panel bei einem oder mehreren Attributen niedrige F-Werte auf, so wie in Abbildung 10 dargestellt, bedeutet dies, dass die Panellisten für diese Attribute noch weitere Schulungen benötigen.

5.3.2.2 Wiederholbarkeit

Die **Wiederholpräzision** zeigt, wie bereits beschrieben, die „Übereinstimmung wiederholter Messungen einer Probe zu gleichen Bedingungen wie gleiche Prüfer, gleicher Ort und gleiche Zeit der Sitzung“. Dies lässt sich statistisch durch die **ANOVA** (Analysis of variance) mittels der Produkt-Wiederholung-Interaktion bewerten. Weiterhin stellen Diagramme der **mittleren Fehlerquadrat-Summe (MSE-Werte)** einer einfaktoriellen ANOVA die Fähigkeit der einzelnen Prüfer, reproduzierbare Ergebnisse zu liefern, dar (Abbildung 11). Je kleiner der MSE-Wert ist, desto geringer weicht die Bewertung des Prüfers von Messung zu Messung ab. In der Abbildung 11 sind generell die MSE-Werte auf niedrigem Niveau; Ausnahmen bilden das Attribut F (Hedley, Bart, Jenny) und das Attribut C (Clarence, Jenny).

Die **MSE-Werte** der Prüfer sollten immer in Verbindung mit den **F Plots** interpretiert werden, da geringe MSE-Werte auch durch die gleiche (nicht diskriminierende) Bewertung der Proben verursacht werden kann. Zur Beurteilung der **Vergleichspräzision** (Übereinstimmung von zwei oder mehreren Beurteilungen, die an der gleichen Probe zu unterschiedlichen Bedingungen vorgenommen wurden) wird die Standardabweichung der Panelmittelwerte der unter Vergleichsbedingungen durchgeführten Sitzungen miteinander verglichen.

Aus den bisherigen Ausführungen ist deutlich geworden, dass sich die Leistungsfähigkeit eines deskriptiven Panels nicht mit einem Parameter bestimmen lässt, sondern dass dazu verschiedene Betrachtungen anzustellen sind. Neben den dargestellten Grafiken bietet PanelCheck eine Funktion, um die Unterscheidungsfähigkeit (P plot) und die Wiederholungsfähigkeit (MSE plot) der Panellisten in einer Grafik darzustellen. Hierzu dient der **p*MSE plot** in Abbildung 12. Es ist hier wünschenswert, dass die p-Werte und MSE-Werte jeweils niedrig sind und möglichst nahe der unteren linken Ecke in der Grafik zu finden sind. Das heißt, dass gemäß Abbildung 12 die Panellisten Jenny und Clarence eine nicht optimale Unterscheidungsfähigkeit und Wiederholpräzision zeigen.

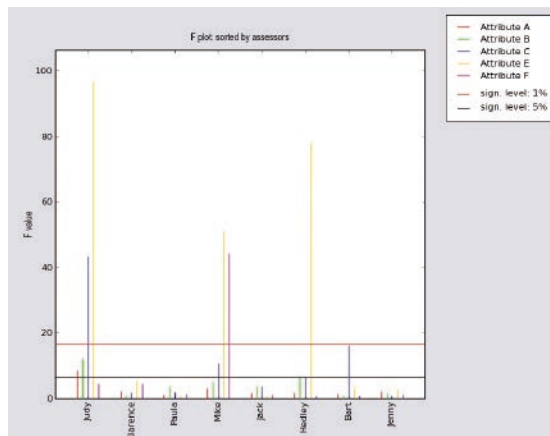


Abbildung 10: F plot (Auswertung aus PanelCheck)

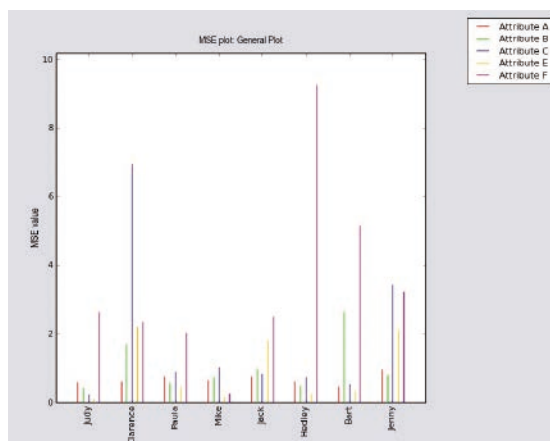


Abbildung 11: MSE plot (Auswertung aus PanelCheck)

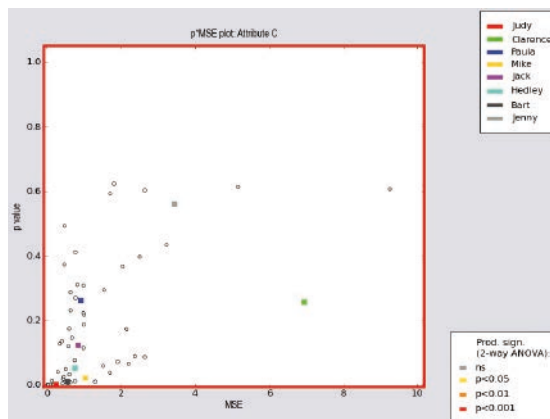


Abbildung 12: p^* MSE plot = gemeinsame Darstellung von F und MSE plot (Auswertung aus PanelCheck)

5.3.2.3 Homogenität

Die **Homogenität** der Bewertungen ist auch ein wesentlicher Punkt. Liegt eine signifikante Produkt-Prüfer-Interaktion vor, beurteilen die Prüfer die Produkte signifikant unterschiedlich oder nutzen die Skalen sehr different, zeigt sich im Ergebnis demzufolge keine Paneleinigung. Auch dies lässt sich über das Software-Tool relativ einfach auswerten. In verschiedenen Profile plots, in denen die Daten der Prüfpersonen dargestellt sind, kann man Wechselwirkungen und Bereiche der Skalennutzung für jede Prüfperson ablesen. Der Profile plot in Abbildung 13 liefert Details zu den Ergebnissen der Prüfpersonen. In diesem Beispiel fallen insbesondere die Panelisten Mike und Paula auf, da sie von den anderen Panelisten abweichen.

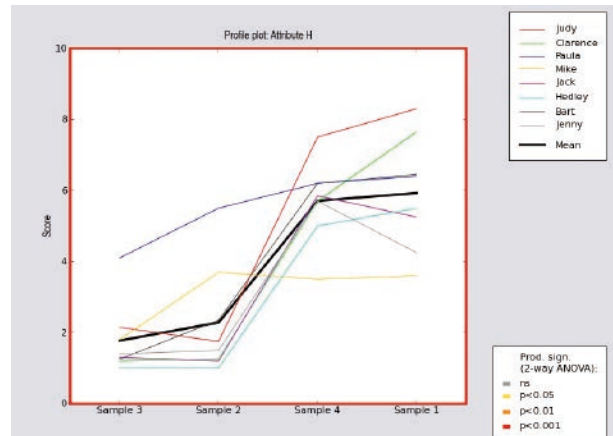


Abbildung 13: Profile plot (Auswertung aus PanelCheck)

Weitere Software-Tools sind im Anhang zu dieser Publikation aufgelistet. Sofern im eigenen Unternehmen entsprechende Softwareprogramme angeschafft werden sollen, empfiehlt sich die direkte Kontaktaufnahme mit den Anbietern und eine anschließende innerbetriebliche Test- und Schulungsphase die jeweilige Software betreffend.

6. Zusammenfassung und Ausblick

Teil 3 der Publikationsreihe „Praxisleitfaden zur Panelschulung“ legt die große Bedeutung des Prüfer- und Panelmonitorings dar. Humansensoriker, deren Arbeitsweise vergleichbar mit chemisch-physikalischen Messgeräten ist, bedürfen vor ihrem Einsatz als Prüfinstrument einer zielgerichteten Kalibrierung. Die Anforderungen der „Guten Laborpraxis“ gelten auch für die „Gute Sensorikpraxis“. Auf Basis eines definierten, auf den Anwendungs- und Einsatzbereich fokussierten Anforderungsprofils an die sensorische Leistungsfähigkeit gilt es, projektbezogenen Prüfpersonen über eine Vorauswahl zu selektieren und über Schulungen gezielt zu trainieren. Die sowohl während der Trainingsphasen als auch in Projekten gewonnenen Prüfergebnisse und -daten sind die Basis für die fortlaufende Überwachung der Leistungsfähigkeit der Humansensoriker im Hinblick auf Unterscheidungsfähigkeit, Wiederholbarkeit und Homogenität. Pannelleitern und Projektleitern im Bereich Sensorik sollen die zuvor beschriebenen Darstellungen als Bestätigung für ihr bereits implementiertes Prüfer- und Panelmonitoring dienen bzw. Motivation und Anregungen für fortlaufende Qualitätsverbesserungen in diesem für den zuverlässigen Einsatz sensorisch-analytischer Prüfer und Panels elementaren Prozess geben. Wenngleich ein umfassendes, gut durchdachtes Prüfer- und Panelmonitoring auf den ersten Blick aufwendig erscheint, lohnt sich der konzeptionelle Aufwand im Vorfeld. Im fortlaufenden Prozess amortisiert sich dieser Input und erhält durch die zunehmende Digitalisierung sämtlicher Daten eine optimale Transparenz und damit eine weitere Arbeitserleichterung.

Literatur:

- Busch-Stockfisch, M. (Hrsg.): Praxishandbuch Sensorik in der Produktentwicklung und Qualitätssicherung. Hamburg: Behr's Verlag, 2002 (44. Ergänzungslieferung 2019)
- Carpenter, R.; Lyon, D.; Hasdell, T.: Guidelines for Sensory Analysis in Food Product Development and Quality Control. 2nd Ed. Gaithersburg: Aspen Publ., 2000
- DLG-Ausschuss Sensorik: DLG-Fachvokabular Sensorik. Frankfurt/M: DLG-Verlag, 2015
- Norm DIN EN ISO 8586:2014-05 – Allgemeiner Leitfaden für die Auswahl, Schulung und Überprüfung ausgewählter Prüfer und Sensoriker
- Kemp, S.; Hollowood, T.; Hort, J.: Sensory Evaluation: A practical handbook. Chichester: Wiley-Blackwell, 2009
- Lawless, H.; Heymann, H.: Sensory Evaluation of Food. 2nd Ed. New York: Springer, 2010
- Muñoz, A.; Civille, G.; Carr, T.: Sensory evaluation in quality control. New York: Van Nostrand Reinhold, 1992
- Norm DIN 10976:2016-08 – Sensorische Prüfung: Difference from Control-Test (DfC-Test)
- Norm DIN EN ISO 5492:2008-10 – Sensorische Analyse: Vokabular
- Norm DIN EN ISO 8586:2014-05 – Allgemeiner Leitfaden für die Auswahl, Schulung und Überprüfung ausgewählter Prüfer und Sensoriker
- Norm DIN EN ISO 8589:2014-10 – Allgemeiner Leitfaden für die Gestaltung von Prüfräumen
- Norm DIN EN ISO 11132:2017-10 – Leitlinien zur Überprüfung der Leistungsfähigkeit eines quantitativen sensorischen Panels
- Norm DIN EN ISO 13299:2016-09 – Allgemeiner Leitfaden zur Erstellung eines sensorischen Profils
- Norm DIN ISO 5725-2:2012-12 – Genauigkeit (Richtigkeit und Präzision) von Messverfahren und Messergebnissen – Teil 2: Grundlegende Methode für Ermittlung der Wiederhol- und Vergleichspräzision eines vereinheitlichten Messverfahrens
- Norm DIN EN ISO/IEC 17025:2018-03 – Allgemeine Anforderungen an die Kompetenz von Prüf- und Kalibrierlaboratorien

- PanelCheck: <http://www.panelcheck.com/Home> [Zugriff am 14.01.2020]
- Stone, H.; Sidel, J.: Sensory Evaluation Practices. San Diego: Academic Press, 1985
- Tomic, O.; Luciano, G.; Nilsen, A.; Hyldeg, G.; Lorensen, K.; Næs T.: Analysing sensory panel performance in a proficiency test using the PanelCheck software. European Food Research Technology 230 (2010), S. 497-511
- Tomic, O.; Nilsen, A.; Martens, M.; Næs, T.: Visualization of sensory profiling data for performance monitoring. LWT-Lebensmittelwissenschaft und Technologie 40 (2007), S. 262-269
- Raithatha, C.: Panel performance measures. In: Rogers, L.: Sensory panel Management. Duxford: Woodhead Publishing, 2018
- Raithatha, C.; Rogers, L.: Panel quality management. In: Kemp, S.; Hort, J.; Hollowood, T.: Descriptive Analysis in Sensory Evaluation. Chichester: John Wiley & Sons, 2018

Auswahl an Software mit unterschiedlichen Funktionen zur Auswertung der Leistungsfähigkeit der Prüfpersonen

- Compusense: Programm zur Organisation, Durchführung und Auswertung von sensorischen Untersuchungen. Weitere Information unter <https://www.compusense.com/en/>
- FiZZ: Programm zur Organisation, Durchführung und Auswertung von sensorischen Untersuchungen. Weitere Information unter <https://www.biosystemes.com/en/fizz-software.php>
- PanelCheck: Programm zur Auswertung von sensorischen Untersuchungen. Weitere Information unter <http://www.panelcheck.com/Home/softwarefeatures>
- Red Jade: Programm zur Organisation, Durchführung und Auswertung von sensorischen Untersuchungen. Weitere Information unter <https://redjade.net/sensory-analysis-software/>
- SenPaq: Programm zur Auswertung von sensorischen Untersuchungen. Weitere Information unter <https://www.qistatistics.co.uk/product-category/software/>
- SensomineR: Programm zur Auswertung von sensorischen Untersuchungen, das auf der Statistiksoftware R basiert. Weitere Information unter <http://sensominer.free.fr/>
- XLSTAT: Programm zur Auswertung von sensorischen Untersuchungen, das in die Tabellenkalkulation Microsoft Excel eingebunden wird. Weitere Information unter <https://www.xlstat.com/de/>

Die Übersicht erhebt keinen Anspruch auf Vollständigkeit und stellt auch keine Empfehlung dar.

Autoren:

Projektteam des DLG-Ausschusses Sensorik; www.DLG.org/sensorikausschuss.html

Unter Federführung von:

- Prof. Dr. Jörg Meier, Hochschule Neubrandenburg, FB Agrarwirtschaft und Lebensmittelwissenschaften, Neubrandenburg, jmeier@hs-nb.de
- Annette Bongartz, ZHAW, Institut für Lebensmittel- und Getränkeinnovation, Forschungsgruppe Lebensmittel-Sensorik, Wädenswil, Schweiz, annette.bongartz@zhaw.ch
- Dr. Jeannette Nuessli Guth, ETH Zürich, Department für Gesundheitswissenschaften und Technologie, Zürich, Schweiz, jnuessli@ethz.ch
- Prof. Dr. Rainer Jung, HS Geisenheim University, Institut für Oenologie, Geisenheim, rainer.jung@hs-gm.de
- Dr.in Eva Derndorfer, Wien, eva@derndorfer.at
- Bianca Schneider-Häder, DLG-Fachzentrum Lebensmittel, Frankfurt/M., B.Schneider@DLG.org

Kontakt:

DLG-Fachzentrum Lebensmittel, Bianca Schneider-Häder, sensorik@DLG.org

© 2020

Alle Informationen und Hinweise ohne jede Gewähr und Haftung. Vervielfältigung und Übertragung einzelner Textabschnitte, Zeichnungen oder Bilder – auch für den Zweck der Unterrichtsgestaltung – nur nach vorheriger Genehmigung durch DLG e.V., Marketing, Eschborner Landstraße 122, 60489 Frankfurt am Main.



DLG e.V.

Fachzentrum Lebensmittel

Eschborner Landstraße 122 · 60489 Frankfurt am Main

Tel. +49 69 24788-311 · Fax +49 69 24788-8311

FachzentrumLM@DLG.org · www.DLG.org