

Effiziente Planung, Vorbereitung, Durchführung und Auswertung sensorischer Prüfungen (Teil 2)



Im Teil 1 des DLG-Expertenwissens zum Thema „Effiziente Planung, Vorbereitung, Durchführung und Auswertung sensorischer Prüfungen“ **Ausgabe X/2012** wurden bereits die Zieldefinition und verschiedene zu berücksichtigende Aspekte zur Testplanung vorgestellt. Der Teil 2 dieses DLG-Arbeitsblattes führt die noch offenen Punkte rund um das Test-Design fort und fokussiert darüber hinaus die Testdurchführung und die Testauswertung.

2.6 Prüfprotokoll

Die Gestaltung des Prüfprotokolls ist ebenfalls wichtig im Zusammenhang mit der Planung eines sensorischen Tests. Es sollte Angaben wie den Namen des Prüfers, Datum und Uhrzeit des Tests, die Bezeichnung des Prüfgutes und die Code-Nr. der Probe/n enthalten. Auch Aussagen zum Gesundheitszustand des Prüfers sind hilfreich, um im Nachhinein ggf. nachvollziehen zu können, warum ein Prüfer abweichend zu allen anderen geurteilt hat.

Kernstück des Protokolls ist die genaue Prüfanweisung mit der zu beantwortenden Fragestellung. Hierbei ist darauf zu achten, dass präzise und für jedes Panelmitglied verständlich formuliert wird. Fehlerhafte und missverständliche Formulierungen können einen sensorischen Test vollkommen ruinieren.

Schließlich muss das Protokoll noch die Möglichkeit bieten, das Prüfergebnis einzutragen oder anzukreuzen. Auch hier sollte man darauf achten, dass unmissverständlich klar ist, wie dies erfolgen soll. Zum Beispiel muss man bei einer Rangordnungsprüfung ausschließen, dass die Prüfer die wahrgenommene Rangfolge der Proben verkehrt herum notieren. Hilfreich ist auch ein zusätzliches Feld „Bemerkungen“. Hier sollten die Prüfer z. B. notieren, wenn Ihnen während des Tests Unstimmigkeiten aufgefallen sind wie Inhomogenitäten in einer Probe, die zu Problemen bei der Bewertung geführt haben, unterschiedliche Probenemperaturen oder eine zu geringe Probenmenge. Bei einer Unterschiedsprüfung können sie zudem anmerken, ob sie bezüglich des Ergebnisses sicher sind oder dieses lediglich vermutet oder gar geraten haben. Dies kann eine wichtige Zusatzinformation sein. Ein Protokollbeispiel zeigt Abbildung 4.

Abbildung 4: Prüfprotokoll für eine paarweise Vergleichsprüfung

Name der Prüfperson:		Datum:
Prüfmuster: Merkmal/ Merkmalseigenschaft:		Gesundheitszustand:
Prüfanweisung: Prüfen Sie die Proben auf dem Tablett in der vom Prüfungsleiter angegebenen Weise von links nach rechts. Rückkosten ist erlaubt. Welche der beiden Proben ist in Bezug auf das genannte Merkmal bzw. die Merkmalseigenschaft stärker ausgeprägt? Falls Sie nicht sicher sind, geben Sie die Probe an, bei der Sie am ehesten vermuten, dass es die stärker ausgeprägte ist. Notfalls müssen Sie raten.		
Prüfproben (Code-Nummern eintragen)	stärker ausgeprägte Probe (Code-Nummer eintragen)	richtig/falsch
.../...	...	
Bemerkungen:		
<input type="checkbox"/> sicher erkannt geraten	<input type="checkbox"/> vermutet	<input type="checkbox"/>
Sonstige Bemerkungen:		

2.7 Prüfzeitpunkt und Ablaufregime

Auch der Prüfzeitpunkt will wohlüberlegt sein. Ungünstig ist es, sensorische Prüfungen unmittelbar nach einer Mahlzeit durchzuführen, weil dann die Prüfer gesättigt und wenig geneigt sind, weitere Lebensmittel zu sich zu nehmen. Die Prüfer sollten aber auch nicht hungrig sein. Prüfzeiten mitten am Vor- oder Nachmittag eignen sich daher am besten. Bei der Prüfung alkoholischer Getränke sollten Zeiten am Nachmittag sicherheitshalber vermieden werden, da ansonsten nicht garantiert werden kann, dass der Blutalkoholspiegel bis zum Feierabend wieder im zulässigen Bereich liegt. Kraftfahrer könnten ansonsten in Bedrängnis kommen.

Bei manchen Lebensmittelprüfungen ist ein bestimmtes Ablaufregime einzuhalten. So werden intensiver schmeckende Produkte, alkoholreichere Getränke oder Proben mit anhaltendem Nachgeschmack in der Regel am Ende einer Sitzung geprüft.

2.8 Testanzahl je Sitzung

Die Testanzahl je Sitzung sollte nicht zu groß sein. Eine hohe Anzahl zu prüfender Proben bzw. Probensätze fördert die Adaption und damit eine Abnahme der Empfindlichkeit der Sinne. Sie führt außerdem zur Ermüdung der Prüfer. Darüber hinaus lassen sich verschiedene Proben immer weniger auseinanderhalten; ihre sensorischen Abbilder im Gedächtnis überlagern sich zunehmend. Diese Effekte sind umso stärker ausgeprägt, je ähnlicher sich die Proben sind.

Die Anzahl durchzuführender Tests je Sitzung hängt auch von der gewählten Prüfmethode ab. So lassen sich z. B. mittels paarweiser Vergleichsprüfung oder „A“-„Nicht A“-Test deutlich mehr Probenvergleiche in einer Sitzung durchführen als mit einem Dreieckstest, da bei letzterem viel mehr hin- und her gekostet werden muss, um zu einem Ergebnis zu gelangen.

Auch die Beschaffenheit der Proben ist bei der Testanzahl je Sitzung zu berücksichtigen. Je intensiver im Geschmack die zu prüfenden Proben sind, je höher ihr Alkoholgehalt ist oder je langanhaltender der Nachgeschmack, desto weniger Tests je Sitzung sind möglich.

2.9 Neutralisation

Im Rahmen der Prüfplanung sind auch Überlegungen bezüglich der Neutralisation zwischen den einzelnen Proben anzustellen. Durch Neutralisation sollen Reste der vorhergehenden Probe ausgespült und die Adaption der Geschmackszellen rückgängig gemacht werden. Andererseits fördert ein Neutralisieren zwischen den Proben das Vergessen wie die vorherige Probe geschmeckt hat. Insbesondere bei Unterschiedstests, wo sich sehr ähnliche Proben in einem Proben-Set befinden, ist es daher manchmal besser, die Adaption in Kauf zu nehmen und auf das Neutralisieren innerhalb des Probensets zu verzichten. Vor einer Wiederholung des Tests zur Absicherung des Ergebnisses oder vor dem Übergang auf ein weiteres Proben-Set ist jedoch in jedem Fall zu neutralisieren.

Zu überlegen ist auch, welches Neutralisationsmittel verwendet werden soll. Für viele Lebensmittel sind ein mineralstoffarmes, weitgehend geschmacksneutrales stilles Mineral- oder Leitungswasser sowie Weißbrot gut geeignet. Für fettreiche Lebensmittel wie Margarine oder Schokolade ist (kaltes) Wasser hingegen weniger zu empfehlen. Da warmes Wasser von vielen nicht gern getrunken wird, kann man stattdessen einen sehr schwach aufgebrühten, nur kurze Zeit gezogenen, unaromatisierten schwarzen Tee verwenden. Er darf keinesfalls adstringierend (die Mundschleimhaut zusammenziehend) und nicht zu heiß sein. Bei scharf schmeckenden Lebensmitteln eignen sich Milch und ggf. Weißbrot, in jedem Fall gefolgt von Wasser, zur Neutralisation.

2.10 „Warm-up“

Günstig ist es, vor einer sensorischen Prüfung ein so genanntes „Warm-up“ durchzuführen. Dazu reicht man den Prüfern in anonymisierter Form, allerdings ohne Codierung, entweder die später im Test zu prüfenden Proben oder eine ähnliche Probe. Dies dient dem Zweck, dass sich die Prüfer bereits an die Probe/n gewöhnen können. Im Falle bewertender Prüfungen (z. B. DLG) findet anhand einer solchen Einstimmungsprobe noch einmal ein Abgleich des Qualitätsmaßstabs der Prüfer statt. Vor dem eigentlichen Test werden diese Einstimmungsproben wieder eingesammelt und es wird unbedingt neutralisiert.

2.11 Kontakt mit den Prüfern

Bei Verbrauchertests muss entschieden werden, ob man die Prüfer selbständig testen lässt, indem man ihnen die Proben und das Prüfprotokoll überlässt, oder ob eine beauftragte Person den Test im Kontakt mit den Prüfern durchführt, diesen die Proben reicht, die Fragestellung formuliert und die Ergebnisse protokolliert. Die erste Methode ist weniger zeit- und personalaufwändig. Allerdings besteht die Möglichkeit, dass es durch Missverständnisse bezüglich der Fragestellung des Tests, durch Verwechslung der Proben oder durch Fehler beim Ausfüllen des Ergebnisprotokolls zu einer Verfälschung des Ergebnisses kommt. Die zweite Methode ist sehr aufwändig, bietet aber die Möglichkeit, derartige Fehler zu vermeiden. Zudem liefert sie mehr Informationen, da auch verbale Äußerungen der Probanden sowie nonverbale Signale mit registriert werden können, was im Hinblick auf das Testziel, insbesondere wenn es um Beliebtheit / Akzeptanz oder Präferenz von Lebensmittelproben geht, von Bedeutung ist. Andererseits besteht bei dieser Methode die Gefahr, die Prüfer ungewollt zu beeinflussen. Einen Ausweg stellt eine so genannte Doppelblind-Prozedur dar, bei der weder die Prüfer noch die mit der unmittelbaren Testdurchführung beauftragten Personen das Prüfziel bzw. das Interesse des Auftraggebers kennen.

3. Testdurchführung

Probleme bei der Durchführung sensorischer Tests, die ggf. eine Verfälschung der Prüfergebnisse nach sich ziehen können, stehen im Zusammenhang mit

- Phänomenen der Sinnesphysiologie und -psychologie
- dem Verhalten bzw. Reaktionen der Probanden
- Veränderungen der Proben während des Tests
- Veränderungen der Prüfumgebung während des Tests

3.1 Sinnesphysiologie bzw. -psychologie

Zu den Vorgängen der Sinnesphysiologie bzw. -psychologie, die Einfluss auf das Testergebnis haben können, gehört das so genannte „Carry-over“ (deutsch: Übertragung). Zum einen kann es bei Verzicht auf das Neutralisieren zwischen den Proben oder bei unzureichender Neutralisation dazu kommen, dass sich Proben im Mund vermischen. Zum anderen führt die zuvor getestete Probe zu einer Adaption der Sinneszellen, wodurch diese bei nachfolgenden Tests weniger empfindlich reagieren. In beiden Fällen hilft ausreichendes Neutralisieren, was jedoch das Vergessen sensorischer Eindrücke begünstigt.

Ein „Carry-over“ kann jedoch auch in dem Sinne zustande kommen, dass eine zuvor getestete Probe den Bewertungsmaßstab des Prüfers beeinflusst. So würde eine zuvor verkostete sehr intensive oder qualitativ herausragende Probe eine nachfolgende mittelmäßige Probe weniger intensiv bzw. qualitativ schlechter erscheinen lassen als wenn letztere allein verkostet würde. Nach einer Probe sehr schwacher Intensität oder sehr schlechter Qualität würde dieselbe mittelmäßige Probe hingegen als intensiver bzw. qualitativ besser eingestuft werden. Derartigen Kontrasteffekten kann man bei seriell-monadisch ausgeführten Tests durch ein Vorsortieren der Proben mittels Rangordnungsprüfung oder durch Anwendung eines nachbar-balancierten Prüfplans (siehe dazu z. B. DIN 10967, Teil 4) begegnen.

Auf die Gefahr des Vergessens sensorischer Eindrücke während eines Tests wurde bereits hingewiesen, ebenso auf die Überlagerung der sensorischen Eindrücke im Gehirn des Prüfers.

3.2 Probanden

Einflüsse seitens der Probanden auf das Prüfergebnis können z. B. durch Interaktionen der Prüfer miteinander während des Tests resultieren. Dem kann durch Prüfkabinen und deren geeignete Anordnung entgegen gewirkt werden (Abbildung 5). Auf die mögliche Beeinflussung der Prüfer durch den Prüfungsleiter und andere Kontexteffekte wurde im Zusammenhang mit der Prüfplanung bereits eingegangen, desgleichen auf Veränderungen in der Leistungsfähigkeit

Abbildung 5: Prüfkabinen für sensorische Prüfungen



Quelle: SAM Sensory and Marketing International

der Prüfer während des Test (Trainingseffekt, Ermüdung, nachlassende Motivation). Darüber hinaus kann es während eines Tests zu einer Veränderung der kognitiven Strategie kommen, d. h. der Art und Weise, mit der das Gehirn an die Lösung der Testaufgabe herangeht.

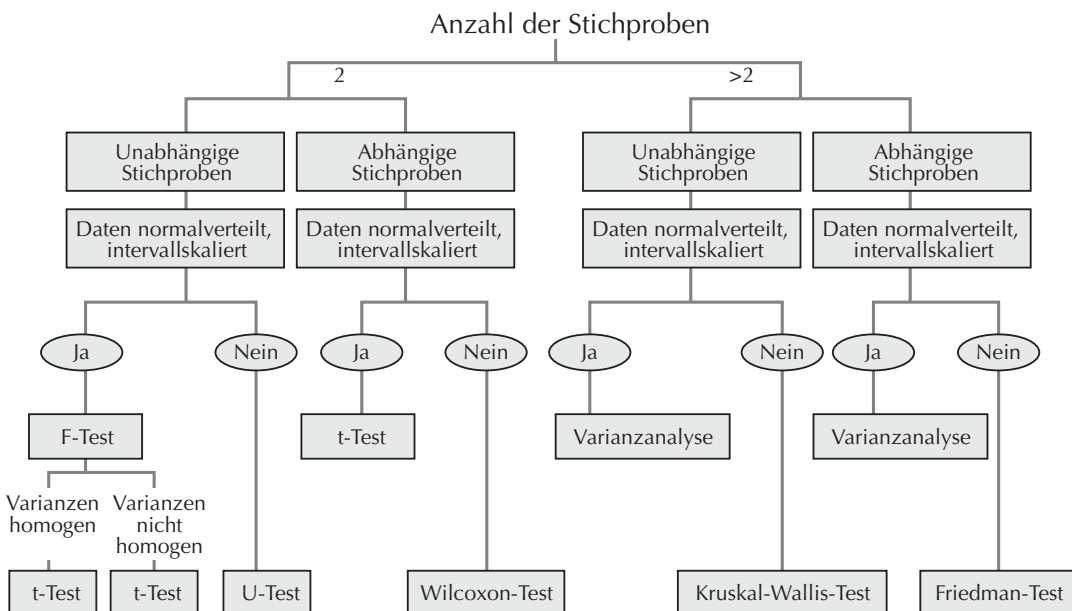
So würde bei einem Dreieckstest zunächst nach der – in welchen Eigenschaften auch immer – abweichenden Probe gesucht. Sobald der Prüfer jedoch glaubt, einen Unterschied in einer konkreten Eigenschaft zu bemerken, konzentriert er sich nur noch auf diese. Derartige Veränderungen lassen sich von außen nicht feststellen und können nur durch Befragen der Prüfer herausgefunden werden.

3.3 Veränderungen der Proben und der Prüfumgebung

Veränderungen der Proben während der Testsitzung können z. B. die Proben temperatur und -textur betreffen. Ebenso kann es zur Verflüchtigung von Aromastoffen kommen.

Auch die Prüfumgebung (Temperatur, Helligkeit, Geruchsbelastung) kann sich während der Sitzung verändern und die Prüfergebnisse beeinflussen.

Abbildung 6: Auswertung von Intensitätsprüfungen



Derartigen Veränderungen kann durch rasches Prüfen, nicht zu lange Sitzungen sowie durch geeignete technische Maßnahmen (Probentemperierung und -abdeckung, Beleuchtung, Be- und Entlüftung sowie Klimatisierung des Prüfraumes) vorgebeugt werden.

Die potenziellen Einflüsse, die während einer sensorischen Prüfung das Prüfergebnis u. U. verfälschen könnten, sollten zuvor gründlich bedacht und im Rahmen der Konzeption des Prüfraumes, der Testplanung bzw. während der Testdurchführung berücksichtigt werden, um verlässliche Ergebnisse zu erhalten.

4. Testauswertung

Bei der Auswertung sensorischer Tests spielen statistische Verfahren eine große Rolle. Welches Verfahren angewendet wird, hängt u. a. von der Natur der Daten und vom Ziel des durchgeführten Tests ab. Daneben ist zu beachten, ob eine einzelne Stichprobe charakterisiert oder zwei bzw. mehrere Stichproben miteinander verglichen werden sollen und ob die Stichproben voneinander abhängig oder unabhängig sind (Abbildung 6).

4.1 Natur der Daten

Hinsichtlich der Natur der Daten unterscheidet man Nominal- bzw. -Kategoriedaten wie sie bei Unterschiedstests anfallen (Anzahl richtiger bzw. falscher Antworten), Ordinaldaten (Ränge), Intervalldaten und Verhältnisdaten. Nur die letzten beiden Datengruppen, die aus der Anwendung von Skalen resultieren, sind ausreichend numerisch im mathematischen Sinne, so dass man mit ihnen rechnen kann (z. B. Berechnung von Mittelwerten, Standardabweichungen, Korrelationen). Beide entstammen einer Skala, bei der die Skalenwerte gleiche Abstände aufweisen. Intervalldaten unterscheiden sich von Verhältnisdaten dadurch, dass die verwendete Skala keinen echten Nullpunkt aufweist.

4.2 Stichproben

Je nachdem, ob eine einzelne Lebensmittelstichprobe charakterisiert werden soll oder ob zwei oder mehrere solcher Proben miteinander verglichen werden sollen, kommen unterschiedliche statistische Methoden zum Einsatz. Bei mehreren Stichproben hat man es in der Sensorik meist mit abhängigen Stichproben zu tun, d. h. die Daten stammen vom gleichen Panel bzw. vom gleichen Prüfer. Dieselben Prüfer haben beispielsweise sowohl eine Probe des Standardprodukts als auch eine Probe mit der neuen, verbesserten Rezeptur geprüft. Unabhängige Stichproben liegen vor, wenn die Proben von verschiedenen Panels getestet wurden, z. B. wenn ein Verbrauchertest parallel in verschiedenen Städten oder zu verschiedenen Zeitpunkten durchgeführt wurde.

4.3 Testziel

Je nach Ziel des sensorischen Tests unterscheidet sich auch die statistische Auswertung. Bei **Unterschiedstest** geht es z. B. darum, herauszufinden, ob ein Unterschied zwischen zwei Proben signifikant („bedeutsam“) ist. Dies ist dann der Fall, wenn die Wahrscheinlichkeit, dass die erhaltenen Prüfergebnisse rein zufällig einen Unterschied widerspiegeln, obwohl dieser nicht existiert, ausreichend gering ist. Auf diese Weise kann man z. B. überprüfen, ob eine vorgenommene Rezepturänderung auch tatsächlich den gewünschten sensorischen Effekt gebracht hat.

Ebenso kann statistisch untersucht werden, ob die geprüften Proben einander signifikant ähnlich sind, ob beispielsweise der Austausch einer teureren durch eine preiswertere Zutat im Endprodukt zum annähernd selben sensorischen Ergebnis führt. In diesem Fall muss die Wahrscheinlichkeit, aufgrund der Prüfergebnisse zu der Schlussfolgerung zu gelangen, dass kein Unterschied existiert, obwohl ein solcher vorliegt, ausreichend gering sein. Bei Rangordnungsprüfungen wird auf die Signifikanz einer gefundenen Rangfolge getestet.

Auch bei **Intensitätsprüfungen** anhand einer Skala, wie sie bei einer Profilanalyse durchgeführt werden, interessiert, ob zwischen Proben signifikante Unterschiede in einem oder mehreren sensorischen Merkmalen bestehen. Daneben kann man aus den Intensitätsmittelwerten ablesen, wie groß die Intensität des jeweiligen Merkmals ist. Die Standardabweichung ist ein Maß für die Streuung der Einzelwerte um den Mittelwert. Mittelwerte und Standardabweichungen werden berechnet, um eine oder mehrere Stichproben zu charakterisieren (beschreibende Statistik) bzw. um anhand dieser Stichproben auf die Beschaffenheit der Grundgesamtheit (z. B. Produktionscharge) zu schließen (folgernde Statistik). Derartige Berechnungen (sog. „parametrische Statistik“) sind jedoch nur zulässig, wenn die Daten normalverteilt sind. Dies ist daher vorab zu überprüfen. Auch normalverteilt erscheinende Intensitätsdaten müssen nicht zwangsläufig normalverteilt sein, da beim Test auf Normalverteilung von gleichen Skalenabständen ausgegangen wird, in den Köpfen der Prüfer die Abstände zwischen den einzelnen Skalenpunkten jedoch in der Realität recht unterschiedlich ausfallen können. So scheuen sich beispielsweise manche Prüfer, sehr hohe oder sehr niedrige Skalenwerte zu vergeben. Dieses Problem bleibt auch bei Verwendung unstrukturierter Linienskalen bestehen. Auch Mittelwerten in der Nähe der Skalenenden kann keine Normalverteilung der Einzelwerte zugrunde liegen, da die Skala nach der einen Seite „keinen Platz“ für eine Normalverteilung bietet. Bei offensichtlich nicht normalverteilten Daten verbieten sich parametrische Auswertungsverfahren. In solchen Fällen ist auf nicht-parametrische statistische Verfahren (z. B. Wilcoxon-Test, Friedman-Test) auszuweichen, auch wenn diese am

Ende nur die Information liefern, ob Unterschiede signifikant sind oder nicht und welche Probe/n intensiver bzw. weniger intensiv im jeweiligen sensorischen Merkmal ist/sind. Eine Aussage über die Größe der sensorischen Unterschiede zwischen den Proben ist nicht möglich. Bei annähernder oder scheinbarer Normalverteilung ist es sinnvoll, die nichtparametrische Auswertung zumindest zur Kontrolle anzuwenden, da nicht bekannt ist, wie sehr die Bedingungen für die parametrische Statistik verletzt sind und welcher Fehler daraus resultiert.

Sofern verschiedene Proben vom selben Panel geprüft wurden, kann man in der Regel davon ausgehen, dass die erhaltenen Intensitätsdaten (abhängige Stichproben) bei allen Proben in etwa gleich streuen (homogene Varianz). Stammen die Intensitätsdaten der Proben jedoch von verschiedenen Panels (unabhängige Stichproben), ist dies nicht zulässig. In dem Fall sind die Daten auf Homogenität der Varianzen zu prüfen (F-Test). Die Homogenität bzw. Nichthomogenität der Varianzen bei unabhängigen Stichproben ist bei den Berechnungen zur Feststellung signifikanter Probenunterschiede (t-Test) zu berücksichtigen.

Eine Übersicht über die statistischen Methoden zur Auswertung von Intensitätsprüfungen zeigt Abbildung 6.

Bei **Akzeptanztests** mit Verbrauchern erhält man ebenfalls Daten, die scheinbar einer Intervallskala entstammen. Hier darf man jedoch noch weniger als bei Intensitätsprüfungen davon ausgehen, dass die Abstände auf der hedonischen Skale in den Köpfen der Verbraucher gleich sind. Daher sind nichtparametrische Auswertungen verlässlicher.

4.4 Kommerzielle Sensorikprogramme

Die Verfügbarkeit kommerzieller Sensoriksoftware mit der Möglichkeit, anspruchsvolle statistische Analysen der in sensorischen Tests gewonnenen Daten durchzuführen und diese grafisch zu veranschaulichen, stellt in der Praxis eine große Erleichterung dar. Es darf jedoch nicht vergessen werden, dass die Ergebnisse der statistischen Auswertung immer nur so verlässlich sein können wie die ihr zugrunde liegenden sensorischen Daten. Einer sorgfältigen Planung und Durchführung der sensorischen Tests, dazu gehört auch die gründliche Schulung eines analytischen Panels, kommt somit allererste Priorität zu. Anderenfalls nützt auch die anspruchsvollste Statistik wenig.

Weiterführende Literatur:

- Busch-Stockfisch, M. (Hrsg.): Praxishandbuch Sensorik in der Produktentwicklung und Qualitätssicherung. Loseblattsammlung. Behr's Verlag, Hamburg
- O'Mahony, M.: Sensory Evaluation of Food. Marcel Dekker Inc., New York und Basel 1986
- Stone, H., Sidel, J. L.: Sensory Evaluation Practices. Academic Press Inc., San Diego, New York, 2004
- Derndorfer, E.: Lebensmittelsensorik. Facultas-Verlag, Wien 2006
- Quadt, A., Schönberger, S., Schwarz, M.: Statistische Auswertungen in der Sensorik. Behr's Verlag, Hamburg, 2009
- Liptay-Reuter, I., Ptach, C.: Sensorische Methoden und ihre statistische Auswertung. Dexheim, 1998
- DIN- bzw. DIN EN ISO-Normen zur Sensorik
- Weitere DLG-Arbeitsblätter Sensorik, insbesondere 02/2010, 04/2010, 05/2010, 01/2011, 06/2011

Autor:

Prof. Dr. Dietlind Hanrieder, Hochschule Anhalt, Bernburg, d.hanrieder@loel.hs-anhalt.de

Kontakt:

Bianca Schneider-Häder, DLG-Fachzentrum Ernährungswirtschaft, Sensorik@DLG.org

In Zusammenarbeit mit dem DLG-Ausschuss Sensorik (www.DLG.org/Sensorikausschuss.html).

© 2012

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., Servicebereich Kommunikation, Eschborner Landstraße 122, 60489 Frankfurt am Main.



DLG e.V.
Fachzentrum Ernährungswirtschaft
Eschborner Landstr. 122
60489 Frankfurt a. M.
Tel.: +49 69 24788-311
Fax: +49 69 24788-8311
FachzentrumLM@DLG.org
www.DLG.org