

Kritische Betrachtungen zur «Bewertenden Prüfung mit Skale» als einem wesentlichen Verfahren der sensorischen Analyse

K. Paulus*, R. Zacharias*, L. Robinson** und H. Geidel***

1.0 Einleitung

Die Verfahren der sensorischen Analyse können (unter anderem) in

- Unterschiedsprüfungen
 - Beschreibende Prüfungen
 - Bewertende Prüfungen
- eingeteilt werden (1).

Bei den Bewertenden Prüfungen erfolgt die Beurteilung von Produktmerkmalen durch einen Zahlenwert (Punkte, Noten) aus einer Skale zum Zweck der Feststellung der sensorischen Qualität. Damit unterscheidet sich die Bewertende Prüfung grundsätzlich von anderen Prüfverfahren, bei denen ebenfalls Skalen verwendet werden, zum Beispiel Intensitätsprüfung oder Hedonische Prüfung. Die Intensitätsprüfung als solche ist nicht bewertend. Die Skale gibt lediglich den Zahlenwert für die Intensität an, mit der ein bestimmtes Merkmal wahrgenommen werden kann, ohne dass hiermit eine Wertung vorgenommen wird. Bei der Hedonischen Prüfung wird der Grad der Bevorzugung eines Produktes bzw. die Einstellung des Konsumenten zu einem Produkt mit Hilfe einer verbalen, graphischen oder numerischen Skale festgehalten. Die direkte Erfassung einzelner Merkmalseigenschaften eines Produktes und ihre Ausprägung ist Grundlage der (beschreibenden) Profilanalyse.

Die Aufgabe der Bewertenden Prüfung mit Skale ist eine möglichst umfassende Charakterisierung der Probe eines Produktes auf der Basis der Beurteilung von mehreren Merkmalen. Die Skale soll dabei die üblicherweise auftretenden typischen bzw. fehlerhaften Ausprägungen eines Merkmals berücksichtigen.

«Merkmale» sind zum Beispiel Farbe, Form, Geruch, Geschmack, Textur, «Merkmalseigenschaften» sind Teilkomponenten der Merkmale, wie zum Beispiel Süße beim Geschmack oder Saftigkeit bei der Textur. Unter Ausprägung ist die Intensität von Merkmalseigenschaften zu verstehen, die sowohl positiv als auch negativ zur Bewertung eines Merkmals beitragen können (vergl. auch 3.2). Ein vorliegender Normentwurf über die Bewer-

tende Prüfung mit Skale beinhaltet lediglich grundsätzliche Aussagen in bezug auf Anwendbarkeit, Durchführung und Auswertung dieses Verfahrens (2).

Für die Aufstellung von Skalen und die Differenzierung in einzelne Merkmale ist eine Norm in Vorbereitung. Die historische Entwicklung der Bewertenden Prüfung sowie mögliche Modifikationen sind bereits ausführlich behandelt worden (3). Mit der verstärkten Anwendung dieses Prüfverfahrens in Forschung und Praxis traten jedoch auch damit zusammenhängende Unklarheiten und Probleme stärker zutage. Ziel dieser Arbeit ist es daher, einige diesbezügliche Aspekte zu diskutieren.

2.0 Das Verfahren

Zunächst sollen in allgemeinerer Form Aufbau und Anwendungsmöglichkeiten der Bewertenden Prüfung mit Skale erläutert werden.

2.1 Aufbau einer Bewertenden Prüfung mit Skale

Eine Bewertende Prüfung mit Skale kann als ein zweidimensionales System bezeichnet werden. Die eine Koordinate umfasst die bei der Charakterisierung der Produktqualität berücksichtigten Produktmerkmale, die andere Koordinate stellt den zur Beurteilung dieser Merkmale zugrundegelegten Masstab, die Skale, dar (Abb. 1).

		Produktmerkmale										
		A	B	C	D	E	•	•	•	•	N	
Skale	n											
	n-1											
	n-2											
	•											
	•											
	1											

Abb. 1 Prinzipieller Aufbau einer Bewertenden Prüfung mit Skale.

*Bundesforschungsanstalt für Ernährung, D-7500 Karlsruhe

**Institut für Lebensmitteltechnologie und Verpackung e.V., D-8000 München

***Universität Hohenheim, D-7000 Stuttgart-Hohenheim

Tab. 1 Allgemeines Schema der Bewertenden Prüfung mit Skale (nach [3])

Merkmale	1. Qualitätsklasse erwünschte Merkmalsausprägungen 9 8 7	2. Qualitätsklasse tolerierbare Merkmalsausprägungen 6 5 4	3. Qualitätsklasse unerwünschte Merkmalsausprägungen 3 2 1
Farbe	Besonders ansprechende, leuchtende, natürliche Farbe, die für die Probe typisch ist.	Leuchtende, natürliche Farbe, die für die Probe typisch ist, etwas zu blass oder zu dunkel, wenige Exemplare schwach verfärbt erlaubt.	Starke Verfärbung der Probenoberfläche, z.B.: bläulich oder graustichig. Vollständige Verfärbung der Probenoberfläche in abstossende Farböne. Farbe nicht mehr wahrnehmbar.
Form	Vollkommene Formhaltung, pralle, unbeschädigte Form.	Sehr gut erhaltene Form, einzelne leicht veränderte Exemplare.	Formen im allgemeinen stark eingefallen, geschrumpft. Stark veränderte Formen, Noch nicht abtossend, Fortgeschrittene Auflösung.
Geruch	Besonders feiner ausgeprägter, artspezifischer Geruch.	Voller, kräftiger, artspezifischer Geruch.	Veränderter Geruch, z.B.: vollkommen abgestanden, leicht vergoren. Nicht mehr ansprechend. Unangenehm veränderter Geruch. Noch nicht abtossend, z.B.: ranzig, fischig, stark heutig.
Geschmack	Besonders feiner ausgeprägter, artspezifischer Geschmack.	Voller, kräftiger, artspezifischer Geschmack.	Veränderter Geschmack, z.B.: vollkommen abgestanden, leicht ranzig oder vergoren. Nicht ansprechend. Unangenehm veränderter Geschmack. Noch nicht abtossend, z.B.: ranzig, fischig, stark heutig.
Textur	Besonders gute, artspezifische Textur, z.B.: festes, saftiges Gewebe, sehr zart.	Sehr gute artspezifische Textur, z.B.: festes Gewebe, zart.	Veränderte Textur, z.B.: sehr unehentlich, zu weich, zu hart, leicht ledrig oder zäh. Unangenehm, veränderte Textur, z.B.: vollkommen breilig, ausgesprochen hart.

Tab. 2 Spezielles Bewertungsschema für hitzesterilisierte Kartoffeln (nach [3])

Merkmale	1. Qualitätsklasse erwünschte Merkmalsausprägungen			2. Qualitätsklasse tolerierbare Merkmalsausprägungen			3. Qualitätsklasse unerwünschte Merkmalsausprägungen		
	9	8	7	6	5	4	3	2	1
Farbe	sortentypisch leuchtend völlig einheitlich fleckenlos	sortentypisch kraftig einheitlich fleckenlos	sortentypisch überwiegend einheitlich fleckenlos	sortentypisch leicht uneinheitlich, hellere oder dunklere Stellen, leicht stumpf	leicht dunkel, blass, matt, stumpf, teilweise uneinheitlich, leicht fleckig	Verfärbungen (z.B. nach weiss, grau, braun) stark uneinheitlich, deutlich fleckig	deutliche Verfärbungen, völlig uneinheitlich, stark fleckig	starke Verfärbungen	völlig verfärbt
Form / Formerhaltung	völlig glatte Oberfläche, unbeschädigte Form	glatte Oberfläche, unbeschädigte Form	leicht aufgeraute Oberfläche, unbeschädigte Form	Oberfläche aufgeraut bis krümelig, unbeschädigte Form, vereinzelt leichte Risse	Oberfläche stark krümelig, durch Risse leicht beschädigte Form	Oberflächenteile abgelöst, starke Rissbildung und teilweise Auflockerung der Form in grosse Stücke	Form deutlich beschädigt, in grosse Stücke zerfallen nicht mehr erkennbar	ursprüngliche Form durch zu starkes Zerfallen	breiartig
Geruch	rein ausgeprägt harmonisch	rein abgerundet	rein, leicht abgeflacht, nicht mehr ganz abgerundet	nicht mehr ganz rein, abgeflacht, leicht herb, leicht streng, leicht erdig	flach, herb, streng, leicht dumpf, leicht brandig, leicht säuerlich, erdig	sehr flach, dumpf, brandig, süss, seifig, stark erdig, säuerlich	völlig leer, stark dumpf, stark brandig, stark süss, stark seifig, sauer	völlig fremd	völlig fremd, abtossend
Flavour (Gesamtgeschmack)	rein ausgeprägt harmonisch	rein abgerundet	rein, leicht abgeflacht, nicht mehr ganz abgerundet	nicht mehr ganz rein, abgeflacht, leicht herb, leicht streng, leicht erdig	flach, herb, streng, leicht dumpf, leicht brandig, leicht seifig, säuerlich, erdig, leichter Nachgeschmack	sehr flach, dumpf, brandig, süss, seifig, erdig, futterkartoffelartig, säuerlicher Nachgeschmack	völlig leer, stark dumpf, stark brandig, stark süss, stark seifig, sauer, starker unangenehmer anhaltender Nachgeschmack	völlig fremd	völlig fremd, abtossend
Textur	fest, leicht feucht, feinkörnig, einheitlich	fest, leicht feucht, feinkörnig, überwiegend einheitlich	fest, leicht hart, leicht weich, leicht wässrig, leicht trocken, feinkörnig, überwiegend einheitlich	teilweise etwas fest, zu hart, zu weich, zu trocken, zu wässrig, leicht grobkörnig, leicht pappig, leicht speckig, leicht uneinheitlich	insgesamt etwas zu hart, zu weich zu trocken, zu wässrig, leicht grobkörnig bis griesig, pappig, leicht zäh, uneinheitlich	hart, weich, trockenen, feuchtnass, griesig, stark klebrig, stark speckig, zäh, stark uneinheitlich	sehr hart, sehr weich bis musig, sehr trocken, sehr feucht, stark griesig, völlig klebrig, speckig, zäh, stark uneinheitlich	völlig uneinheitlich bis verändert	völlig verändert

Bei der Bewertenden Prüfung mit Skale ist davon auszugehen, dass nur eine Skale gleicher Punktzahl für alle zu bewertenden Merkmale angewendet wird. Zur Frage des Skalenumfanges sowie der Auswahl der Merkmale, die die Produktqualität charakterisieren, wird später noch Stellung genommen.

Basis des Verfahrens ist ein unspezifisches Bewertungsschema, in dem in Form allgemeiner Beschreibungen eine Zuordnung zwischen Punkten und Merkmalseigenschaften erfolgt (z.B. das in **Tab. 1** aufgeführte Schema). Hierbei entsprechen vereinbarungsgemäss ansteigende Zahlenwerte ansteigender Merkmalsqualität.

Die Erarbeitung spezieller Bewertungsschemata (z.B. für ein bestimmtes Produkt) basiert auf der Aufstellung von Ausdrucksspektren zur Charakterisierung der speziellen Gegebenheiten. Hierfür sind die zur Beschreibung der Eigenschaften festgelegten Ausdrücke, also die Merkmalsdifferenzierungen, soweit wie möglich aufzufächern und den einzelnen Notenbereichen zuzuordnen. Ein Beispiel für ein solches spezielles Bewertungsschema ist in **Tab. 2** wiedergegeben.

Bei der Abstufung der für die Charakterisierung des Merkmals herangezogenen Eigenschaften müssen zwei Punkte besonders berücksichtigt werden. Zunächst besteht die Aufgabe, die maximale Qualität eines Produktes festzulegen. Je nach Fragestellung kann dies die jeweilige «Frischware» oder das jeweilige bestmögliche oder standardisierte be- oder verarbeitete Produkt sein. So wird es in vielen Fällen nicht möglich sein, ein sterilisiertes Produkt beispielsweise mit einem aus der selben Rohware frisch hergestellten direkt zu vergleichen. Insofern ist es bei der Erstellung eines solchen Schemas unerlässlich, die obere Qualitätsgrenze festzulegen, da sonst die eine Grenze des Systems fließt und damit die ganzen Zuordnungen bzw. letztlich die Bewertungen unsicher und schwammig ausfallen. Auch bei aus Einzelkomponenten zusammengesetzten Produkten (z.B. Fertiggerichte) gibt es i.a. keinen direkten Vergleich mit der frisch zubereiteten Rohware. Somit muss ein spezielles Bewertungsschema die spezielle Situation, gegeben durch die Produktart bzw. die Problemstellung an einem bestimmten Produkt berücksichtigen.

Der zweite wesentliche Punkt zur Standardisierung des Qualitätsspektrums eines Merkmals ist die Wortbeschreibung. In diese Standardisierung muss eine an der Verbrauchererwartung orientierte Qualitätsvorstellung einfließen. Demnach sollte der eigentlichen Erstellung des Bewertungsschemas jeweils ein Verbrauchertest zur Festlegung der Anforderungen vorgeschaltet sein, um zunächst grob festzustellen, was unter einer hohen, mittleren und geringen Qualität zu verstehen ist. Die auf dieser Grobeinteilung basierenden Transformationen zwischen Ausprägungen der Merkmale einerseits und Punkten bzw. Noten andererseits müssen dann von Sachkennern festgelegt und entsprechend definiert werden.

2.2 Anwendungsmöglichkeiten

Die Bewertende Prüfung mit Skale wird angewendet, um die sensorische Qualität oder bestimmte Merkmale der Produkte zwischen Erzeugung und Verbrauch zu bewerten. Auf den speziellen Normentwurf wurde bereits hingewiesen (2). Die Anwendungsmöglichkeiten der sensorischen Prüfverfahren sind bereits in einer entsprechenden Norm zusammengefasst (4).

Prinzipiell handelt es sich bei der Bewertenden Prüfung mit Skale um eine Einzelprobenprüfung, wobei je nach Aufgabenstellung ein oder mehrere Produkte nebenein-

ander durch geschulte Prüfpersonen geprüft werden. Die gestellte Aufgabe kann unter anderem sein:

- Produktvergleich im Sinne einer Qualitätskontrolle bzw. -überwachung, z.B. Prüfungen der DLG, der Stiftung Warentest, Einhaltung von Qualitätsstandards
- Bestimmung des Qualitätsniveaus von Produkten, z.B. Prüfung von Konkurrenzmustern industrieller Herstellung
- Festlegung von Richtwerten für die Haltbarkeit langfristig lagerfähiger Produkte, z.B. Bestimmung der Lagerfähigkeit tiefgefrorener Erzeugnisse
- Vergleich der Rohwaren- und Fertigproduktqualität in Abhängigkeit von verfahrenstechnischen Einflussfaktoren, z.B. Einfluss von Art und Zusammensetzung der Rohware, Verarbeitung, Behandlung, Verpackung und Lagerung
- Grobe Vorsortierung mehrerer Proben für eine daran anschließende weitere spezielle sensorische Prüfung, z.B. Feststellung von Unterschieden einzelner, spezifischer Merkmalseigenschaften von 2 Proben, Ermittlung der Bevorzugung zum Zweck der Marktforschung

Eine weitere Unterteilung der Anwendungsmöglichkeiten ergibt sich nach der Anzahl der Prüfmuster, da die Auswertung der in Zahlenwerten vorliegenden Prüfurteile danach zu erfolgen hat.

2.2.1 Prüfung von einem Prüfmuster. Diese Prüfung wird dann notwendig, wenn Proben ohne ursächlichen Zusammenhang zu bewerten sind. Dies ist z.B. bei der Qualitätskontrolle, der Durchführung von Prämierungen oder der Beurteilung von Konkurrenzmustern der Fall, da hier sowohl die Rohware als auch die Herstellungs- und evtl. auch Lagerungsbedingungen unterschiedlich sind. Weiterhin gehört hierzu die Untersuchung z.B. des Lagerverhaltens eines bestimmten Prüfmusters mit dem Ziel der Festlegung des Haltbarkeitsdatums. Darüber hinaus wird eine Prüfung mit einem Prüfmuster immer erforderlich sein, wenn aufgrund von Produkteigenschaften eine Prüfung mehrerer Produkte nebeneinander nicht möglich ist. So sind beispielsweise Proben mit langanhaltendem Nachgeschmack nur in zeitlich grösserem Abstand beurteilbar.

2.2.2 Prüfung von zwei Prüfmustern. Voraussetzung für diese Prüfung ist, dass entweder bei gleicher Ausgangsware ein Parameter (z.B. Rezeptur, Temperatur, Zeit, Verpackung) zwei Varianten aufweist oder umgekehrt zwei unterschiedliche Produkte z.B. bei der gleichen Temperatur oder in einer Verpackungsart in gewissen Zeitabständen zu beurteilen sind. Die Problemstellung ist somit in jedem Fall die Frage, ob sich Unterschiede in der Gesamtqualität oder einzelner Qualitätsmerkmale zwischen zwei Prüfmustern ergeben bzw. ob sich zwei Prüfmuster in eine bestimmte, mit Zahlenwerten definierte Rangordnung zueinander bringen lassen.

2.2.3 Prüfung von mehr als zwei Prüfmustern. Sind von mehr als zwei Prüfmustern einzelne Merkmale nebeneinander oder die Gesamtqualität zu bewerten, können sich folgende Anwendungsmöglichkeiten ergeben:

- Ein Ausgangsprodukt und ein Einflussfaktor mit drei oder mehr Varianten, z.B. unterschiedliche Lagertemperaturen, Verpackungsarten, Rezepturen bzw. Zubereitungsverfahren
- Ein Ausgangsprodukt und mehr als zwei Einflussfaktoren mit einer oder mehreren Varianten, z.B. unter-

schiedliche bzw. gleiche Lagertemperaturen, Verpackungsarten und Rezepturen

- Zwei und mehr Ausgangsprodukte und mehr als zwei Einflussfaktoren mit einer oder mehreren Varianten, z.B. verschiedene Rohwaren (Sorte, Herkunft) und Herstellungs- und Lagerbedingungen

Die Begrenzung der vorgelegten Probenzahl ist dabei von den versuchstechnischen Möglichkeiten, von der Anzahl der zu beurteilenden Merkmale und vor allem von der Qualifikation bzw. dem Ermüdungsgrad der Prüferpersonen abhängig. Je nach Schwierigkeitsgrad und Genauigkeitsanforderung sollte bei diesen nebeneinander erfolgenden Prüfungen eine Anzahl von sechs Proben nicht überschritten werden. Die Problemstellung ist auch hier wie bei der Prüfung von zwei Prüfmustern, Unterschiede zwischen den Prüfmustern mit Hilfe der Skale zu ermitteln, und damit die Proben in eine definierte Rangordnung zueinander zu bringen.

3.0 Diskussion einiger kritischer Punkte der Bewertenden Prüfung mit Skale

Ungeachtet der deutlichen Vorteile, die diese Arbeitsweise im Vergleich zu der Prüfung mit einem unspezifischen Schema mit sich bringt, ist eine Reihe Fragen genereller Art bei der Abfassung bzw. Neuerstellung von Schemata zu berücksichtigen.

3.1 Zusammenhang zwischen Merkmal bzw. Merkmalsausprägung und Skale

Unabhängig vom Skalenumfang, der allgemein ausgedrückt von 1 bis n geht, kommt dem Zusammenhang zwischen Merkmalsausprägung und Skale grösste Bedeutung zu. In **Abb. 2** sind verschiedene Kurven eingezeichnet, die die prinzipielle Diskussion der Zusammenhänge etwas erleichtern sollen.

Kurve 1 repräsentiert einen linearen Zusammenhang zwischen Ausprägung und Skale. Die Linearität ist an sich angestrebt, die Realisierung bereitet jedoch erhebliche Schwierigkeiten, da die physiologischen Vorgänge bei der Erfassung der einzelnen Merkmale und auch die exakten Zusammenhänge zwischen Ausprägung eines Merkmals und physiologischem Reiz nicht umfassend

bekannt sind. Die damit häufig auftretende Streckung bzw. Raffung der Ausprägung eines Merkmales führt dann zu Zusammenhängen, wie sie qualitativ durch die Kurven 2 und 3 dargestellt sind. Daneben sind durchaus Zusammenhänge denkbar, bei denen die Kurve eine bzw. mehrere Wendepunkte aufweist, das heisst Streckung und Raffung wechseln sich innerhalb des gesamten Bereiches ab (vergl. Kurve 4). Die Hauptproblematik liegt demnach eindeutig bei der Skalenzuordnung, also der Transformation der Ausprägung der einzelnen Merkmale in Punkte bzw. Noten.

Die in **Abb. 2** dargestellten Kurven geben 4 mögliche Transformationsarten wieder. In der Praxis kommt es darauf an, die Ausprägung der einzelnen Merkmale so exakt wie möglich zu definieren. Selbst dann bleibt jedoch die Transformation als Schwierigkeit bestehen, da über die Empfindung der einzelnen Merkmale durch die Sinne bzw. deren mathematischer Zusammenhang noch relativ wenig bekannt ist. Erschwerend kommt weiter hinzu, dass in bezug auf die Ausprägungsart zusätzlich zwei qualitativ unterschiedliche Fälle zu unterscheiden sind (**Abb. 3**).

Entsprechende Kurve 1, die vereinfacht die in **Abb. 2** wiedergegebenen Varianten umfassen soll, besteht ein Zusammenhang zwischen der Ausprägung und der Skale derart, dass mit zunehmender Ausprägung eine Zunahme des Punktwertes einhergeht oder umgekehrt (z.B. bei der Ausbildung von Fehlkomponenten). In der Praxis sind hierbei, wie bereits erwähnt, gewisse Abweichungen von diesem linearen Zusammenhang möglich bzw. unvermeidbar, entscheidend ist aber der kontinuierliche Anstieg zwischen minimaler und maximaler Ausprägung. Ein solcher kontinuierlicher Zusammenhang trifft beispielsweise auf die Merkmalseigenschaft «Härte» zu, wobei minimale Ausprägung ein sehr weiches Produkt, maximale Ausprägung ein sehr hartes Produkt darstellen würde.

Im 2. Fall, entsprechend Kurve 2, ergibt sich ein Ausprägungsmaximum, das nicht mit den Endpunkten der Skale zusammenfällt. Es findet eine Entwicklung von einer nicht vorhandenen Ausprägung zur maximalen Ausprägung statt und danach erfolgt wiederum ein Abfall dieser Ausprägung. Ein Beispiel hierfür wäre der Garungsvorgang, wo von einer sehr harten Textur über einen optimalen Zustand eine zu weiche Textur erreicht werden kann.

Es ist also sehr wesentlich, dass bei der Erarbeitung

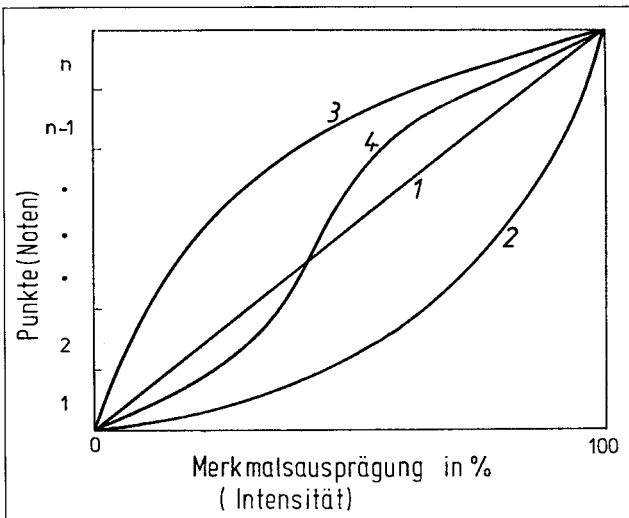


Abb. 2: Prinzipieller Zusammenhang zwischen Merkmalsausprägung (Intensität) und Skale

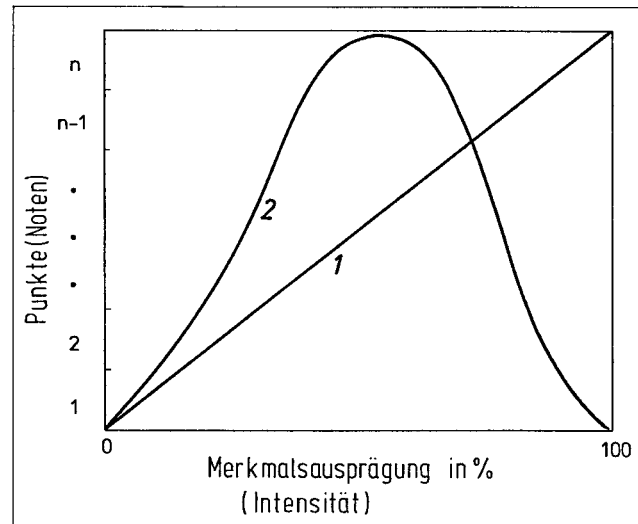


Abb. 3: Qualitative Darstellung der prinzipiellen Möglichkeiten der Ausprägung eines Merkmals

eines Schemas für die Bewertende Prüfung mit Skale diese grundlegenden Zusammenhänge berücksichtigt werden und dem Problem der Skaleneinteilung, also der Transformation der Merkmalsausprägungen in adäquate Notenabstufungen bzw. Notenbereiche grosse Aufmerksamkeit geschenkt wird.

3.2 Zusammenhang zwischen Genusswert und den einzelnen Merkmalen

Zur Erfassung der Produkteigenschaften stehen fünf verschiedene Sinnesindrücke zur Verfügung: Olfaktorische, gustatorische, visuelle, auditorische und haptische Eindrücke (1).

Zu den olfaktorischen Eindrücken gehören Sinneswahrnehmungen, bei deren Zustandekommen keine direkten Empfindungen der Mundhöhle, also keine Geschmacksreize beteiligt sind. Gustatorische Eindrücke sind demnach solche Geschmackswahrnehmungen, bei denen primär olfaktorische Eindrücke ausgeschlossen sind. Visuelle Eindrücke umfassen die optisch wahrnehmbaren Merkmale (z.B. Farbe, Form), die pauschal mit Aussehen bzw. Erscheinungsbild umschrieben werden können. Auditorische Eindrücke umfassen die mit dem Ohr wahrnehmbaren Sinnesindrücke und spielen naturgemäss nur bei speziellen Produkten eine Rolle. Haptische Eindrücke sind statisch erhaltene Eindrücke, beispielsweise von der Produktoberfläche, sowie kinetische Eindrücke, vornehmlich bzgl. des Gefüges. Weiterhin sind hierunter durch Kälte oder Wärme hervorgerufene sowie durch chemisch-physiologische Reize bedingte Eindrücke zu verstehen. Der Begriff «Flavour» umfasst olfaktorische, gustatorische und haptische Eindrücke und ist somit als Oberbegriff zu werten.

Im Einzelfall kommt es darauf an, diese Eindrücke zur Merkmalsbeschreibung heranzuziehen. Zunächst ergeben sich die groben Merkmalsbereiche Aussehen/Erscheinungsbild, Flavour und Textur bzw. Konsistenz. Diese Bereiche lassen sich dann weiter untergliedern wie z.B. das Aussehen in Farbe und Form, Flavour in Geruch und Geschmack – sowie die bei deren Ausbildung beteiligten haptischen Eindrücke –, so dass man insgesamt zu einer mehr oder weniger starken Auffächerung der einzelnen Merkmalsbereiche kommt (**Abb. 4**). Im allgemeinen geht man von den fünf Merkmalen Farbe,

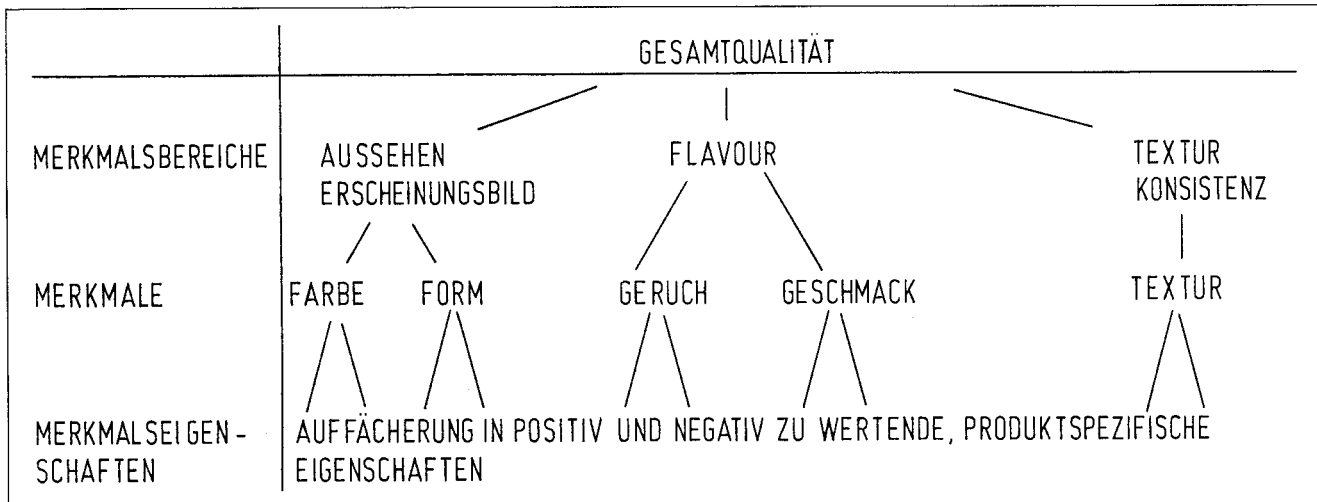
Form, Geruch, Geschmack und Textur aus. Dieser unspezifischen Gliederung ist dann je nach Problemstellung das einzelne Produkt gegenüberzustellen und zu entscheiden, wie ein Produkt am besten durch welche Merkmale charakterisiert werden kann. Von Fall zu Fall kann also in den interessierenden Merkmalsbereichen stärker differenziert, in anderen wiederum zusammengefasst bzw. weggelassen werden. Dieser Merkmalsfestlegung kommt grundlegende Bedeutung zu, da mit den ausgewählten Merkmalen eine umfassende Charakterisierung der sensorischen Produktqualität, also des Genusswertes, bzw. verschiedener Teilqualitäten ermöglicht werden soll. Prinzipiell ist anzustreben, die Produktcharakterisierung mit einem Minimum an hierfür geeigneten Merkmalen zu erreichen. Eine zu starke Aufteilung der Merkmale erscheint nicht sinnvoll; wenn eine solche starke Aufgliederung erforderlich ist, sollte z.B. die Profilanalyse als Verfahren verwendet werden.

Die Produktmerkmale können durch eine unterschiedliche Zahl von Merkmalseigenschaften charakterisiert werden. Diese Differenzierung eines Merkmales wird so vorgenommen, dass sowohl positive als auch negative Eigenschaften berücksichtigt und den Skalenwerten zugeordnet werden. Die Zuordnung orientiert sich an produktspezifischen Qualitätsnormen, die sich wiederum zum Beispiel aus der Verbrauchererwartung und den produktionstechnischen Möglichkeiten ergeben. In **Tab. 1** ist als Beispiel die unspezifische Aufgliederung von Merkmalen aufgeführt, **Tab. 2** stellt ein produktspezifisches Beispiel für eine angestrebte umfassende Beschreibung der sensorischen Qualität dar.

3.3 Umfang der Skale

Die Festlegung des Skalenumfanges n kann zwangsläufig erst nach Klärung der Zusammenhänge zwischen Merkmalsausprägung und zahlenmässiger Erfassung erfolgen. Für die generelle Entscheidung sind zwei Kriterien zu beachten. Zunächst kommt es auf die Aufgabenstellung bzw. auf die wünschenswerte Schärfe des Ergebnisses an. Dies bedeutet jedoch nicht, dass man den Skalenumfang beliebig gross machen kann. Denn n ist durch das zweite Kriterium begrenzt, und zwar durch die Möglichkeit, ein Merkmal hinreichend exakt zu differenzieren. Als Obergrenze hat sich ein Skalenumfang mit n = 9 (Punkte von 9 bis 1) bewährt. Dieser erlaubt z.B. eine symmetrische Dreiteilung der Skale in einem Bereich hoher, mittlerer sowie geringer Qualität. Darüberhinaus bringen einstellige Zahlen Vorteile bei der Auswertung mit sich.

Abb. 4: Zusammenhang zwischen sensorischer Gesamtqualität, Merkmalsbereichen, Merkmalen und Merkmalseigenschaften



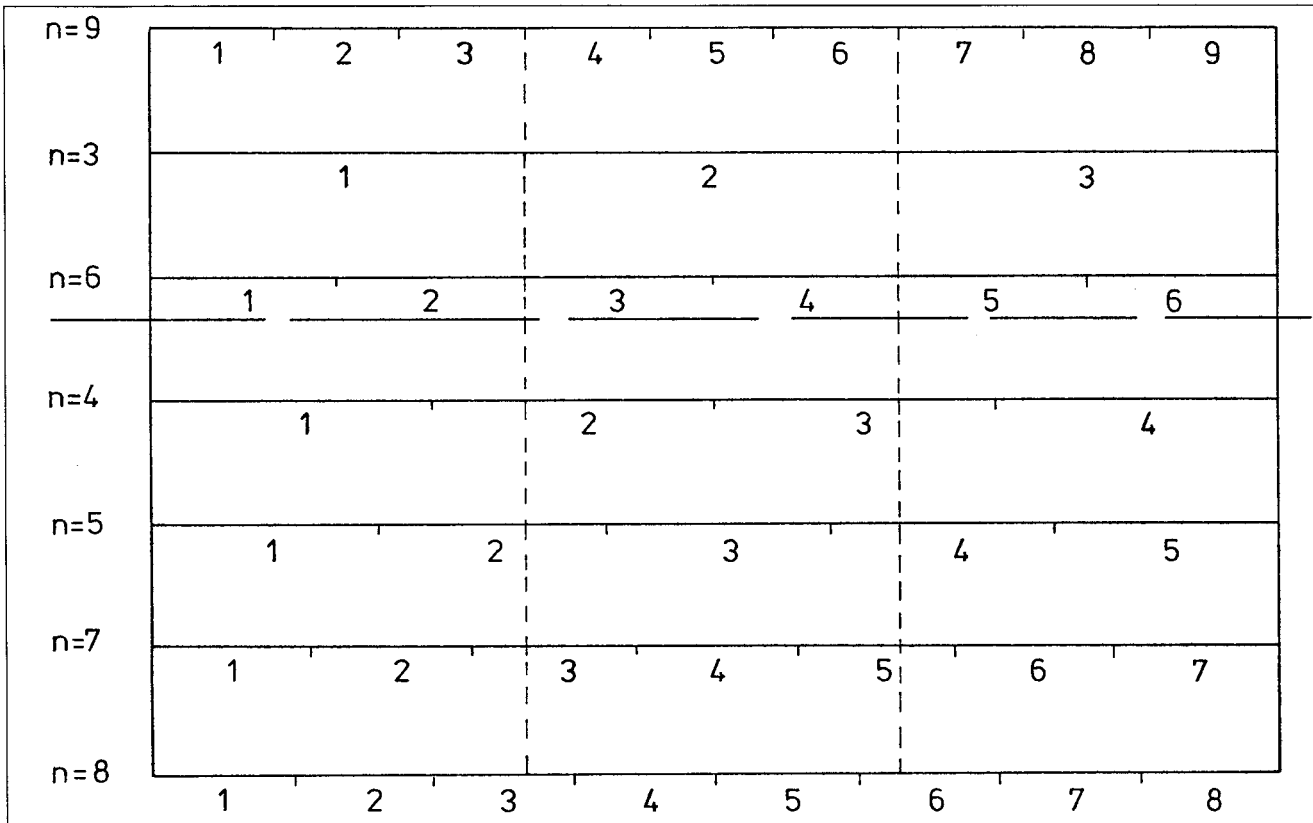
Wenn $n = 9$ als maximal angesehen wird, und wegen der Dreiteilung $n = 3$ nicht unterschritten werden soll, ergibt sich eine Reihe von Möglichkeiten zwischen $n = 4$ bis $n = 8$ (Abb. 5).

Es wird als wenig sinnvoll erachtet, sämtliche Möglichkeiten für n auszunutzen, da dadurch die Übersichtlichkeit stark leidet und auch die Vergleichbarkeit der Ergebnisse verschiedener Arbeitsgruppen erheblich erschwert wird. Als dritte Möglichkeit neben Skalen mit $n = 9$ und $n = 3$ würde sich demnach eine Skale mit $n = 6$ anbieten. Beispiele hierfür gibt es ebenfalls in der Praxis. Somit ergibt sich folgende Abstufung;

- Eine Skale mit $n = 9$ (9 bis 1) entspricht sehr hohen Anforderungen (z.B. für Forschung und Entwicklung).
- Eine Skale mit $n = 6$ (6 bis 1) reicht zur etwas größeren Charakterisierung der Qualität aus.
- Eine Skale mit $n = 3$ (3 bis 1) steht für sehr grobe Aussagen zur Verfügung.

In diesem Zusammenhang ist es wichtig, darauf hinzuweisen, dass die jeweiligen Punkte bzw. Noten nicht einen fixen Punkt, sondern Bereiche wiedergeben, das heisst diese Punkte sind Klassenmitten im mathematischen Sinn. Auch für die praktische Durchführung der Bewertenden Prüfungen mit Skale ist dieser Sachverhalt von Bedeutung. Dadurch, dass die Note tatsächlich einen Bereich angibt, hat der einzelne Prüfer mehr Flexibilität bzw. er muss sich dieser Möglichkeit bewusst werden. Beispielsweise kann die Extremnote n sowohl allerhöchste Ansprüche widerspiegeln als auch an die Grenze zur nächstniedrigen Note $n = 1$ eingereicht werden. Die hier benutzten Punkte bzw. Noten sind im Sinne der Statistik Werte einer Ordinalskale.

Abb. 5: Darstellung der Skalen mit $n = 3$ bis $n = 9$



3.4 Stufenweises Vorgehen unter Verwendung von Skalen mit unterschiedlichem n

Aus verschiedenen Gründen kann es erforderlich sein, dass ein stufenweises Vorgehen bei der Bewertung eines Produktes wünschenswert bzw. notwendig ist. Insbesondere ist dies der Fall, wenn es um die Erarbeitung eines speziellen Bewertungsschemas geht bzw. es sich um die Bewertung eines Produktes handelt, für das ein spezifisches Bewertungsschema noch nicht existiert. Um den Prüfern die Hilfestellung zu geben, wie sie normalerweise durch das spezielle Bewertungsschema, in dem die zugrundezulegenden Qualitätsanforderungen ausgewiesen sind, gegeben wird, erfolgt zunächst eine grobe Bewertung des Produktes bzw. der Produktmerkmale mit einer Skale mit $n = 3$. Auf der Basis dieser größeren Klassifizierung bzw. der angegebenen Urteile ist dann festzulegen, in welchen der drei Bereiche die Qualität des einzelnen Merkmals eingeordnet werden soll. In der danach folgenden 2. Stufe der Bewertung stellen dann die einzelnen Prüfer fest, welcher Stufe die Qualität der einzelnen Merkmale zugeordnet werden muss. In dieser 2. Stufe wird also der jeweilige Bereich wiederum dreigeteilt, d.h. der einzelne Prüfer hat dann die Möglichkeit, innerhalb dieses Bereiches zwischen drei Noten zu wählen. Projiziert auf ein einstufiges Bewertungsverfahren würde sich demnach eine Bewertende Prüfung mit Skale mit $n = 9$ ergeben.

3.5 Bereiche und Grenzwerte sensorischer Qualität

Es gibt bisher keine allgemein gültigen bzw. allgemein akzeptierten Definitionen zur Charakterisierung einer bestimmten sensorischen Qualität. Das Internationale Kälteinstitut (IIF), Paris, hat jedoch für die Frage der Lagerung tiefgefrorener Lebensmittel einige Anhaltspunkte vorgeschlagen und definiert, die durchaus – bei modifizierter Anwendung – allgemeine Gültigkeit haben könnten (5).

Hohe Qualität:

Für den Fall tiefgefrorener Lebensmittel hat das IIF den Begriff «high quality life» so definiert, dass hierunter die Zeit zwischen Gefrieren des Gutes und dem Zeitpunkt zu verstehen ist, nach der mittels einer Unterschiedsprüfung gerade ein signifikanter Unterschied im Hinblick auf die Ausgangsqualität desselben Produktes festgestellt werden kann. Hierbei wird stillschweigend vorausgesetzt, dass die Ausgangsqualität einwandfrei ist. Dieser Begriff hat also nur dann einen Sinn, wenn die sensorische Beurteilung der Produktqualität in die erste Qualitätsklasse fällt. Der erweiterte Begriff «hohe Qualität» gilt somit für die während einer bestimmten Zeit nahezu konstant bleibende Ausgangsqualität eines Produktes. Eine Verallgemeinerung auf Produkte, die unverarbeitet oder in irgendeiner Form be- und verarbeitet sind, erscheint sinnvoll und nützlich (vergl. **Tab. 3**)

Der Bereich «hohe Qualität» würde dann bedeuten, dass die Abweichung von der hohen Ausgangsqualität durch irgendeine Behandlung (z.B. Lagerung) bzw. Behandlungsmodifizierung (z.B. veränderte Rezeptur) mittels einer Unterschiedsprüfung gerade festgestellt werden kann, der effektive Qualitätsunterschied, erfasst mit der bewertenden Prüfung mit Skale, jedoch minimal ist.

Verkehrsqualität: Hiermit soll entsprechend der Definition des IIF («practical storage life») ein Bereich charakterisiert werden, bis zu dessen unterer Grenze die Qualität absinken darf, ohne dass das Produkt hierbei unverkäuflich wird (5). Dieser untere Wert ist bei einer Skale mit $n = 9$ der Note 5 zugeordnet, da davon auszugehen ist, dass zwischen Verkauf des Produktes und Genuss ein weiterer Qualitätsabfall resultiert, oder anders ausgedrückt, dass das Produkt eine weitere Qualitätsreserve beinhaltet. Dies ist der Grund, warum die Verkehrsqualität über der minimalen Verzehrqualität liegen muss. Auch hier wird primär wiederum von Erhaltung einer Qualität (storage life) gesprochen, da es darum geht, einen zeitlichen Ablauf zu definieren bzw. zu fixieren. Verallgemeinert würde Verkehrsqualität bedeuten, dass die Qualität durch irgendeine Behandlung bzw. Behandlungsmodifizierung (z.B. durch Lagerung) bis zur Note 5 absinken darf, wobei die Feststellung der Qualität mit einer Bewertenden Prüfung mit Skale und einem Skalenumfang von neun Noten erfolgt.

Tab. 3 Bereiche und Grenzwerte der sensorischen Qualität

Qualitätsbereich	Notenbereich bei der Skale mit $n = 9$	Grenzwert bei der Skale mit $n = 9$	Erfassbar durch
Hohe Qualität (High Quality Life)	9-7	7	Bewertende Prüfung mit Skale evtl. in Verbindung mit Unterschiedsprüfung
Verkehrsqualität (Practical Storage Life)	9-5	5	Bewertende Prüfung mit Skale
Verzehrqualität (Limit of Acceptability)	9-4	4	Bewertende Prüfung mit Skale
Genussuntauglich (Not Acceptable)	4	-	Bewertende Prüfung mit Skale, auch z.B. bakteriologische Prüfung

Verzehrqualität: Hiermit wird der Wert charakterisiert, bei dem ein Produkt gerade noch für den Verzehr geeignet ist («limit of acceptability»). Verzehrqualität bedeutet also allgemein hier einen Bereich, innerhalb dessen die Qualität durch irgendeine Behandlung bzw. Behandlungsmodifizierung bis zum Grenzwert 4 absinken darf, wobei die Feststellung der Qualität mit einer Bewertenden Prüfung mit Skale und einem Skalenumfang mit neun Noten erfolgt.

Genussuntauglichkeit: Am unteren Ende schliesst sich ein Bereich an, der genussuntaugliche Qualitäten umfasst («not acceptable»). Diese Genussuntauglichkeit kann sowohl durch sensorische Prüfungen ermittelt werden, sofern es die sensorische Qualität ausschliesslich betrifft, jedoch sind auch Fälle denkbar, wo ungeachtet der sensorischen Beurteilung die hygienische Beschaffenheit des Produktes so unbefriedigend ist, dass die Entscheidung über Genussuntauglichkeit auf der Basis der Ergebnisse bakteriologischer Prüfungen zu fällen ist. In **Abb. 6** sind die Bereiche und Grenzwerte für den Fall der Anwendung der Bewertenden Prüfung mit Skale (Skalenumfang 9 Punkte) unter Berücksichtigung früherer Vorschläge zusammengefasst (1, 6-8).

Hieraus wird deutlich, dass eine Dreiteilung einer Skale erhebliche Vorteile mit sich bringt, weshalb, wie in 3.3 ausgeführt, Skalenumfänge mit $n = 3$, $n = 6$ und $n = 9$ ausgewählt werden sollten.

3.6 Probleme bei der statistischen Auswertung

Im Hinblick auf die Beurteilung der Ergebnisse für ein Merkmal sind zunächst drei Fälle zu unterscheiden, wobei davon ausgegangen wird, dass insgesamt i Urteile vorliegen. Diese ergeben sich aus:

- eine Person, i Urteile
- i Personen, je ein Urteil
- p Personen, je q Urteile ($p \times q = i$)

Soweit man nur eine Auswertung der Häufigkeitsverteilungen der abgegebenen Urteile vornimmt, ist es offenbar gleichgültig, wie die i Urteile zustandekommen. Falls man jedoch auch den Einfluss der Prüfer erfassen will, ist es zweckmässig, die letzte Möglichkeit (p Personen mit je q Wiederholungen) zu realisieren. Die Charakterisierung der Beurteilung eines Merkmals erfolgt durch Erstellung von Häufigkeitsverteilungen aus den Einzelnoten und Berechnung des Durchschnittswertes mit Standardabweichung.

Für eine weitere Auswertung ist die Fragestellung von wesentlicher Bedeutung. Die Bewertende Prüfung mit Skale kann

- zur Qualitätsbeurteilung von Einzelproben
- zum Qualitätsvergleich (statisch) oder zur Prüfung der Qualitätsänderung (zeitliche Abhängigkeit)

eingesetzt werden. Im Fall des Vergleiches von Durchschnitts- bzw. Mittelwerten werden Tests zur objektiven Beurteilung der Frage, ob sich zwei Mittelwerte unterscheiden, angewendet (z.B. t-Test). Für den Fall der zeitlichen Abhängigkeit einer Qualitätsänderung kommen z.B. Regressions- und Varianzanalyse in Betracht. Häufig ist auch eine Gesamtnote zur Kennzeichnung eines Produktes erforderlich. Zur Ermittlung der Gesamtqualität bieten sich mehrere Möglichkeiten an:

- Arithmetischer Mittelwert der einzelnen Merkmalsdurchschnittswerte
- Gewogener Mittelwert der einzelnen Merkmalsdurchschnittsnoten unter Beachtung von «Wichtigungen» für die einzelnen Merkmale
- Berechnung eines Mittelwertes nach einer der vorgenannten Möglichkeiten unter Berücksichtigung von «Ausschlussmerkmalen», d.h., soweit bei einem Merkmal eine vorgegebene Grenze unterschritten wird, ist die Gesamtqualität nicht mehr berechenbar
- Der «Mittelwert» entspricht dem Durchschnittswert des schlechtesten Merkmals

Der einzelne Prüfer hat, unabhängig davon, welche der aufgeführten Alternativen gegebenenfalls angewendet wird, nichts mit der Bildung der Note für die Gesamtqualität zu tun. Dies obliegt dem Prüfungsleiter bzw. Sachkennern mit entsprechender Produkterfahrung. Im allgemeinen ist von einer Mittelwertbildung mit Wichtung der Merkmale auszugehen. In dieser Alternative ist auch die oben zuerst genannte enthalten, also dass alle Merkmale mit demselben Wichtungsfaktor belegt werden. Die dritte Alternative, die Berücksichtigung von Ausschlussmerkmalen, beinhaltet dasselbe Vorgehen, wobei jedoch vorher zu prüfen ist, ob alle Merkmale oberhalb eines vorgegebenen Grenzwertes liegen.

Durch die Bildung einer Gesamtnote verwischt sich zwangsläufig das Bild, so daß von dieser Möglichkeit nur dann Gebrauch gemacht werden sollte, wenn keine andere Auswertungsform bzw. Darstellungsart der Ergebnisse in Betracht kommt (z.B. bei der Einordnung eines Produktes in Handelsklassen).

4.0. Zusammenfassung

Die Bewertende Prüfung mit Skale ist ein wesentliches Prüfverfahren der analytischen Sensorik.

Abb. 6 Zusammengefasste Darstellung der Zusammenhänge zwischen Note, Qualitätsbereichen und Grenzwerten der Qualität für die Bewertende Prüfung mit Skale

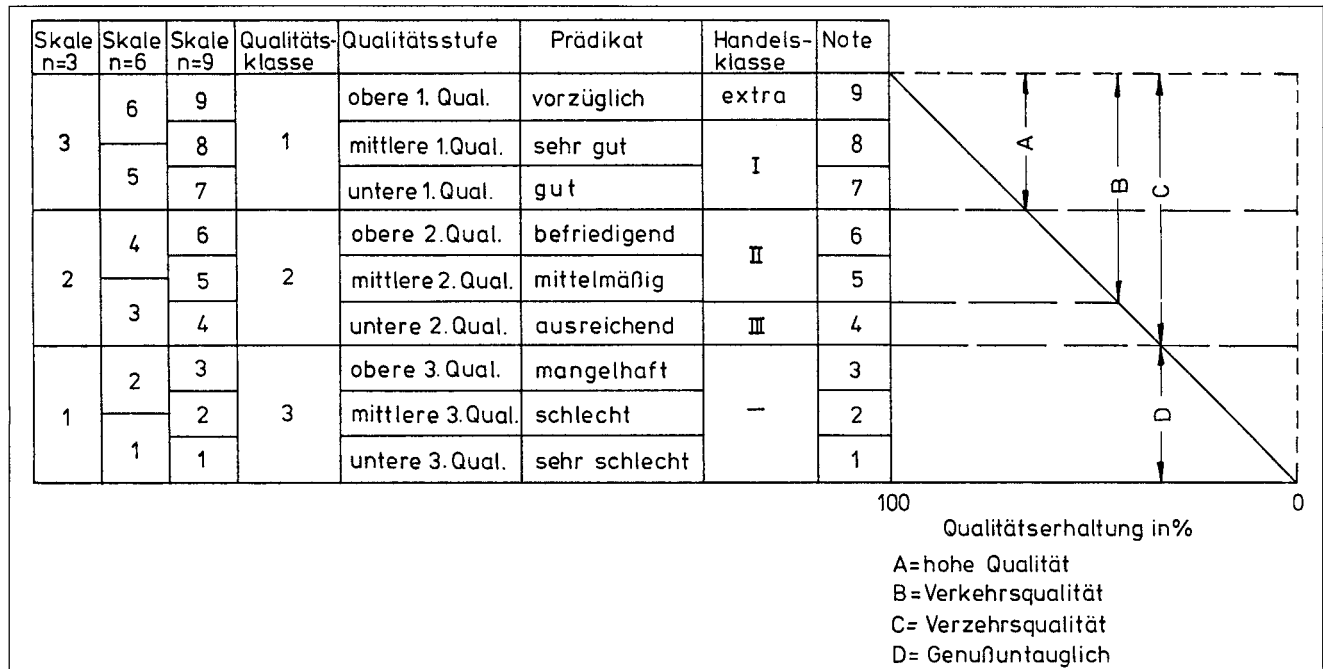
Es wurde versucht, darzustellen, dass auch heute noch der Zusammenhang zwischen Merkmalsausprägung und Skale nicht vollständig geklärt ist, insbesondere unter dem Aspekt der Vielfalt der verschiedenen Produkte und deren einzelner für die Qualität wesentlicher Merkmale. Um so wichtiger erscheint es, die grundlegenden Zusammenhänge zu kennen und sie bei der Aufstellung von Bewertungsschemata zu berücksichtigen. Das Problem der Transformation sensorisch erfassbarer Merkmalsausprägungen in einen Zahlenwert einer Skale kann dadurch erleichtert werden, dass man versucht, die einzelnen Bereiche der Skale so umfassend wie möglich für die einzelnen Merkmale zu charakterisieren. Somit wird die Tendenz immer mehr dahingehen, die Skale selbst durch spezifische Bewertungsschemata zu ergänzen, um die Prüfung zu objektivieren und auch die Ergebnisse, die üblicherweise in Zahlenwerten anfallen, zu präzisieren bzw. zu ergänzen.

Hinsichtlich des Umfanges einer Skale scheinen heute der maximale und minimale Umfang unumstritten zu sein. Es wird für wenig sinnvoll erachtet, auf einen Skalenumfang von $n > 9$ bzw. $n < 3$ zurückzugreifen. Die Frage ist jedoch, ob es sinnvoll ist, die zwischen $n = 3$ und $n = 9$ liegenden Möglichkeiten umfassend auszunutzen. Insofern würde wahrscheinlich eine dritte Skale mit $n = 6$ ausreichen, um den mittleren Genauigkeitsbereich auszufüllen.

Bereiche und Grenzwerte der sensorischen Qualität sind im Hinblick auf die Angaben von Haltbarkeitsdaten, also zur Charakterisierung der Genusstauglichkeit unerlässlich. Für die Bewertende Prüfung mit einer Neun-Noten-Skale werden derartige Bereiche und Grenzwerte vorgeschlagen und diskutiert.

Die statistische Auswertung ist, unabhängig von der der Prüfung zugrundeliegenden Fragestellung, zur Beurteilung der Prüfungsergebnisse und für die Ableitung von weiteren Aussagen unerlässlich. Anwendbare Methoden sind angegeben.

Insgesamt ist festzustellen, dass die Bewertende Prüfung mit Skale zwar ein relativ altes Verfahren der sensorischen Analyse darstellt, jedoch in den letzten Jahren besonders kritischen Betrachtungen unterzogen wurde. Dies hat zweifellos zur Konkretisierung, Präzisierung und auch zur Objektivierung des Verfahrens selbst geführt.



Es erscheint wünschenswert, dass die hier zusammengefassten Ergebnisse einer langjährigen Diskussion zwischen Sachkennern und die hieraus abgeleiteten Vorschläge breite Zustimmung erhalten.

Literatur

- 1 Deutsche Normen: Allgemeine Grundlagen der Sensorik; DIN 10950, 1973, Beuth-Verlag GmbH, Berlin.
- 2 Normentwurf: Sensorische Prüfverfahren; Bewertende Prüfung mit Skale, DIN 10952, Beuth-Verlag GmbH, Berlin.
- 3 PAULUS, K., GUTSCHMIDT, J., FRICKER, A.: Karlsruher Bewertungsschema – Entwicklung, Anwendbarkeit, Modifikationen. *Lebensm.-Wiss.-u.-Technol.* 2 (1969) 132–139.
- 4 Deutsche Normen: Anwendung sensorischer Prüfverfahren; DIN 10953, 1976, Beuth-Verlag GmbH, Berlin.
- 5 International Institut of Refrigeration: Recommendations for the processing and handling of frozen foods; Annexe au Bulletin de l'Institut International du Froid, Paris, 1972.
- 6 Institut für Lebensmitteltechnologie und Verpackung: Qualitätsbeurteilung bei der Verarbeitung und Lagerung von Lebensmitteln; unveröffentlichte Ausarbeitung.
- 7 GUTSCHMIDT, J.: Die Bestimmung der Haltbarkeit von tiefgefrorenen Lebensmitteln, speziell von Fleisch, Fleischwaren und Fleischfertiggerichten. *Die Fleischwirtsch.* 51 (1971) 295–299.
- 8 GUTSCHMIDT, J.: The storage life of frozen chicken with regard to the temperature in the cold chain. *Lebensm.-Wiss. u. -Technol.* 7 (1974) 137–141.