

## The Beauty-Contest-Game

### Wirtschaftsdidaktische Lerndiagnostik mithilfe eines spieltheoretischen Unterrichtsexperimentes

*Prof. Dr. Ewald Mittelstädt \**, *Till Sender \*\**, *Prof. Dr. Andreas Liening \*\**

*\* Professur für Betriebswirtschaftslehre insbesondere Entrepreneurship Education, Fachhochschule Südwestfalen*

*\*\* Lehrstuhl für Wirtschaftswissenschaft und Ökonomische Bildung, Technische Universität Dortmund*

### Zusammenfassung

Das „Beauty-Contest-Game“ nach Keynes ist ein spieltheoretisches Experiment, um durch Zahlenwahl das Verhalten von Akteuren an Finanzmärkten zu simulieren. Im vorliegenden Beitrag wird diskutiert, ob das Beauty-Contest-Game als Unterrichtsexperiment für ein Diagnostikinstrument ökonomischer Bildung geeignet ist. Die Weiterentwicklung von spieltheoretischen Experimenten zu Lerndiagnostikinstrumenten ermöglicht einen genuin wirtschaftsdidaktischen Zugang. Damit wird das vorhandene Repertoire leistungsorientierter Diagnostikinstrumente (Assessment View), wie es Klausuren oder der Wirtschaftskundliche Bildungstest (WBT) darstellen, um einen Prozessfokus erweitert (Treatment View). Die Güte des Beauty-Contest-Game als ökonomisches Diagnostikinstrument wird durch eine auf Triangulation ausgelegte Korrelationsanalyse mit den beiden assessmentorientierten Verfahren Klausur und WBT festgestellt sowie seine Aussagekraft unter Beachtung von Limitationen erörtert.

### Abstract

According to Keynes, the Beauty Contest Game is a game theory experiment simulating players' behaviour on the financial markets by selecting numbers. The following paper discusses whether the Beauty Contest Game is suitable as a teaching experiment for a diagnostic instrument in Economics Education. Further developing game theory experiments into learning diagnostic instruments provides genuine access to Economic Education. In doing so, current existing performance-oriented diagnostic instruments (assessment view), as featured in class tests or the German Wirtschaftskundliche Bildungstest [Economics Education Test] (WBT), are extended to include a focus on process (treatment view). Conducting a correlation analysis interpreted through triangulation, the quality of the Beauty Contest Game as an economics diagnostic instrument is determined within both the assessment-oriented methods of the examination and the WBT. Moreover, its significance considering the limitations shall also be discussed.

## 1 Fachdidaktische Diagnostik zwischen Assessment und Treatment

Das Lehramtsstudium Wirtschaftswissenschaft befähigt u. a. dazu, wirtschaftsdidaktisch analysieren, argumentieren und gestalten zu können. Dies bezieht sich einerseits auf die Intentionen, Inhalte, Methoden und Medien von ökonomischen Bildungsmaßnahmen, andererseits aber auch auf deren Lernbarkeit unter Berücksichtigung von Zielgruppe und Rahmenbedingungen. Erziehungswissenschaftliche Antworten auf die Lernbarkeit sind „inhaltsneutral“, d. h. sie sind auf generelle Probleme und Vorgänge im Unterricht gerichtet. Das Alleinstellungsmerkmal der Fachdidaktik ist es, die inhalts- und die prozessbezogene Sichtweise zu verbinden (Terhart 2002, 31). Dafür ist es notwendig, die unterrichtlichen Leistungsergebnisse, aber auch die fachspezifischen Vorkenntnisse und Aufgabenschwierigkeiten zutreffend einschätzen zu können. Dies ist das Feld fachdidaktischer Lerndiagnostik. Diagnostik wird jedoch häufig nur auf die Notengebung bezogen. Bereits Schrader (1989) differenziert jedoch zwischen einer Diagnostik zur Messung von Leistungsergebnissen und einer Diagnostik bezogen auf die Lernbarkeit. Während die Leistungsdiagnostik auf die Erteilung von Qualifikationen zielt (Assessment View), fokussiert die Lerndiagnostik eine prozessbezogene Unterrichtsoptimierung (Treatment View).

Ein fachdidaktisches Lerndiagnostikinstrument ist besonderen Rahmenbedingungen ausgesetzt: extreme Ressourcenknappheit. Leistungsorientierte Diagnostikinstrumente werden in der Regel zu Beginn und/oder nach Ende einer Bildungsmaßnahme eingesetzt. Aufgrund ihres formalen Charakters (z. B. Aufnahmetests oder Klausuren) steht ausreichend Zeit und Personalkapazität dafür zur Verfügung. Anders verhält es sich mit lernorientierten Diagnostikinstrumenten, die von der Lehrkraft in den Unterrichtsprozess zu integrieren sind. Im Unterrichtsgeschehen wirkt eine Vielzahl von Informationen auf die Lehrkraft ein. Für aufwendige Auswertungen von Lernaufgaben und die Dokumentation von Beobachtungen sowie deren Reflektion vor dem Hintergrund spezifischer Lernendenmerkmale bleibt wenig Zeit (Kramer 2002), zumeist sind keine weiteren Personalkapazitäten dafür vorgesehen. Unter Berücksichtigung dieser Ressourcenknappheit wäre es daher sehr hilfreich, wenn Wirtschaftslehrkräften je nach Inhalt ökonomische Bildungstests zur Verfügung stünden.

Der vorliegende Beitrag zielt darauf ab, ein Instrument vorzustellen, welches im Sinne einer lernorientierten Diagnostik die inhalts- und prozessbezogene Sichtweise auf Wirtschaftsunterricht verbindet und bezogen auf die Handhabbarkeit als Indikator funktioniert. Gerade wenn das Instrument in Diagnostikprozessen zur Anwendung kommt, sind die üblichen Gütekriterien sorgfältig zu untersuchen, damit fundierte Aussagen über das Einsatzspektrum

und die Limitationen des Instrumentes getroffen werden können.

Im Folgenden wird zunächst das Instrument vorgestellt, es handelt sich um das spieltheoretische Unterrichtsexperiment „Beauty-Contest-Game“ nach Keynes. Es wird dargelegt, warum es inhaltlich für die wirtschaftsdidaktische Lerndiagnostik geeignet ist (Konstruktvalidität). Anschließend wird das Untersuchungsdesign entwickelt, auf dessen Grundlage die Güte des Indikators untersucht wurde. Das Untersuchungsdesign ist als Triangulation ausgelegt, mithilfe der Korrelationsanalyse mit zwei leistungsorientierten Verfahren (Klausur und Wirtschaftskundlicher Bildungstest) wird die kriteriumsbezogene, innere Validität des Instruments empirisch untersucht. Der Beitrag schließt mit einer Diskussion von Limitationen und Einsatzmöglichkeiten.

## 2 Spieltheoretisches Unterrichtsexperiment zur Lerndiagnose

### 2.1 The Beauty-Contest-Game

#### 2.1.1 Ursprung

Das nun im Blickpunkt stehende und hier als „Beauty-Contest-Game“ bezeichnete spieltheoretische Experiment geht zurück auf den britischen Ökonomen John Maynard Keynes, der mit folgenden Worten die dem Spiel zugrunde liegenden Charakteristika beschreibt (Keynes 1936, 131):

*„Die professionelle Geldanlage kann mit jenen Zeitungswettbewerben verglichen werden, bei denen die Teilnehmer die sechs hübschesten Gesichter von hundert Lichtbildern auszuwählen haben, wobei der Preis dem Teilnehmer zugesprochen wird, dessen Wahl am nächsten mit der durchschnittlichen Vorliebe aller Teilnehmer übereinstimmt, sodass jeder Teilnehmer nicht diejenigen Gesichter auszuwählen hat, die er selbst am hübschesten findet, sondern jene, von denen er denkt, dass sie am ehesten die Vorliebe der anderen Teilnehmer gewinnen werden, welche alle das Problem vom gleichen Gesichtspunkt aus betrachten.“*

Dieses Gedankenspiel von Keynes, das aufgrund dieser Worte in der Literatur als „Schönheitswettbewerb“ oder „Beauty-Contest-Game“ bekannt ist, beschreibt eine Situation, in der Individuen das Verhalten anderer Individuen antizipieren und in die eigene Entscheidungsfindung mit einbeziehen müssen. Keynes selber betont im obigen Zitat die ökonomische Bedeutung solcher Entscheidungssituationen, denen zum Beispiel Akteure an Finanzmärkten ausgesetzt sind. Um das Verhalten von Individuen in solchen Situationen analysieren zu können, wird die Situation des Schönheitswettbewerbs auf folgende auch als „Zahlenwahlspiel“ bekannte Spielform übertragen. Dabei werden beliebig viele Spieler dazu aufgefordert,

eine Gewinnzahl aus einem vorgegebenen Zahlenintervall (z. B.  $[1,100]$ ) zu wählen. Gewonnen hat diejenige Zahl, die zwei Dritteln des Durchschnitts aller genannten Zahlen am nächsten kommt. Für die eigene Entscheidungsfindung müssen die Teilnehmenden also die Zahlenwahl der anderen Teilnehmenden antizipieren (Nagel 1999).

Die Probanden der Untersuchung wurden genau einer solchen Situation ausgesetzt. Bevor diskutiert wird, ob und wie anhand des Spielverhaltens lerndiagnostische Rückschlüsse gezogen werden können, wird zunächst vorgestellt, welche Verhaltensempfehlung die Spieltheorie für das Beauty-Contest-Game ausgibt. Zudem werden die Befunde bereits durchgeführter Experimente rund um das Beauty-Contest-Game vorgestellt.

### 2.1.2 Spieltheoretische Analyse

Als Teilgebiet der Mikroökonomie untersucht die Spieltheorie Situationen, in denen mehrere Akteure interagieren und leitet rationales Entscheidungsverhalten ab. Zentrales Ergebnis dieser Analyse ist das Nash-Gleichgewicht, welches eine Verhaltensempfehlung bezeichnet, von der ein Abweichen für einen einzelnen Spieler von Nachteil wäre, sofern die anderen Spieler an der Verhaltensempfehlung festhalten. Diese Verhaltensempfehlung führt also zu einer stabilen Spielsituation, in der sich ein einzelner Akteur nicht mehr verbessern kann (Berninghaus et al. 2010, 362).

Die spieltheoretische Analyse des Beauty-Contest-Game ergibt sich aus folgendem Gedankengang: Das Ausgangsintervall ist das Zahlenintervall  $[1,100]$ . Das Maximum dieses Intervalls ist die Zahl 100 und somit ist auch der theoretisch höchstmögliche Durchschnitt aller genannten Zahlen die Zahl 100. Da aber diejenige Zahl gewählt werden soll, die zwei Drittel des Durchschnitts aller genannten Zahlen am nächsten kommt, wäre eine Wahl der Zahlen, die größer als  $\frac{2}{3} \times 100 = 66,6\bar{6}$  sind, irrational. Die Zahlenwahl jedes rationalen Akteurs erfolgt nun aus dem Intervall  $[1, 66,6\bar{6}]$ . Führt man diesen Gedankengang insgesamt  $n$ -mal durch, so erfolgt eine rationale Zahlenwahl aus dem Intervall  $[1, (\frac{2}{3})^n \times 100]$ . Da die Zahl mit  $(\frac{2}{3})^n \times 100$  immer größer werdendem  $n$  sich der Zahl 1 annähert und schließlich unterschreitet, werden sukzessive alle Zahlen bis auf die Zahl 1 eliminiert. Die spieltheoretische Analyse ergibt also, dass die Wahl auf die Zahl 1 fallen sollte. Spielen alle Spieler diese Strategie, würde das Abweichen eines einzelnen Akteurs diesen den (geteilten) Spielsieg kosten. Gemäß obiger Erläuterung handelt es sich bei der Spielsituation, in der alle Spieler die Zahl 1 wählen, um den stabilen und erstrebenswerten Zustand eines Nash-Gleichgewichts (Selten/Nagel 1998, 16).

### 2.1.3 Tatsächliches Spielverhalten im Beauty-Contest-Game

Wie verhalten sich die Teilnehmenden, wenn das Beauty-Contest-Game in der Praxis durchgeführt wird? Diese Frage war bereits Gegenstand einiger Experimente und Untersuchungen. So führten etwa die spanische Zeitschrift *Expansión* und die britische *Financial Times* das Experiment 1997 zeitgleich mit ihrer Leserschaft durch (Bosch-Domènech/Nagel 1997a/b) sowie die Zeitschrift *Spektrum der Wissenschaft* 1998 (Selten/Nagel 1998).<sup>1</sup>

Zunächst ist festzustellen, dass die spieltheoretische Verhaltensempfehlung in keinem der Experimente zum Spielsieg führte – die Gewinnzahlen lagen mit 16,98 (*Expansión*), 12,61 (*Financial Times*) sowie 14,72 (*Spektrum*) deutlich über dem Nash-Gleichgewicht. Selten und Nagel (1998, 17) führen dies darauf zurück, dass die starke Rationalitätsannahme der Spieltheorie in der Praxis offensichtlich nicht erfüllt ist. So konnte den eingeforderten schriftlichen Begründungen der Wahlentscheidung entnommen werden, dass bei Spielern, die eine recht hohe Zahl gewählt hatten, die Wahl willkürlich geschah oder aus nicht den spieltheoretischen Charakter des Spiels berücksichtigenden Beweggründen (z. B. Lieblingszahl, Geburtstag). Die Mehrheit derjenigen, die eine sehr niedrige Zahl ( $< 10$ ) wählten, ließen in ihren Begründungen hingegen erkennen, dass sie die antizipativen, spieltheoretischen Charakteristika des Spiels verstanden. Zudem gehen rationale, das Spiel durchschauende Akteure zwar davon aus, dass einige der anderen Akteure die Spielsituation nicht durchschauen, schätzen die Anzahl dieser Spieler aber als viel zu gering ein. Stattdessen unterstellen sie, dass die Mehrheit auch zu diesen Gedankengängen fähig ist und eine ähnliche Spielweise wie sie an den Tag legen wird. Somit wird eine höhere aber immer noch sehr niedrige Zahl gewählt, da das eigene Verhalten in zu großem Ausmaß auf die Mehrheit übertragen wird, was in der Psychologie als „false consensus effect“ (Ross et al. 1977) bezeichnet wird und eine Verhaltensanomalie darstellt (Mittelstädt 2011, 90). Festzuhalten bleibt aber: Die Spieler, die die interaktive Situation durchschauen und spieltheoretische Gedankengänge entwickeln, wählen eine niedrige Zahl (Nagel/Bosch-Domènech 1999, 10).

### 2.1.4 Zahlenwahl als Indikator ökonomischer Bildung?

Zentrales ökonomisches Bildungsziel ist die Fähigkeit zur eigenständigen Teilhabe am Wirtschaftsleben. Diese Fähigkeit kann selbstverständlich nicht durch eine einzige Zahlenwahl erfasst werden, zumal ökonomische Bildung als theoretisches Konstrukt nicht direkt messbar

---

<sup>1</sup> Beachte: *Financial Times* und *Spektrum der Wissenschaft* führten das Spiel mit dem Intervall  $[0,100]$  statt  $[1,100]$  durch. Entsprechend der zuvor genannten Argumentation ist die spieltheoretische Verhaltensempfehlung dann die Zahl 0.

ist. Die Zahlenwahl im Beauty-Contest-Game stellt zwar eine interaktive Handlungssituation dar, diese ist jedoch abstrakt und nicht lebenssituativ. Zudem wird weder ein Überblick über ökonomische Grundfertigkeiten einzelner Probanden geliefert, noch wird die Breite ökonomisch relevanter Inhalte und Denkmuster erfasst.

Gleichwohl spricht sich auch Retzmann (2008) mit Hinweis auf gewisse Limitationen grundsätzlich für die Aufnahme spieltheoretischer Situationen in ökonomische Bildungsmaßnahmen aus. Die Bewältigung interaktiver Entscheidungssituationen, wie sie das Beauty-Contest-Game erzeugt, weist durchaus Parallelen zu gängigen Herausforderungen in der Wirtschaftspraxis auf – man denke nur an das Aufeinandertreffen verschiedener Akteure auf Märkten, wo die Zusammenführung von Individualentscheidungen in eine Marktentscheidung mündet. Keynes selber macht dies deutlich am Beispiel der Finanzmärkte und Börsenspekulationen, in welchen die Antizipierung der Handlungen von anderen Marktteilnehmenden von besonderer Bedeutung ist. Demnach bestimmen Marktmeinungen und Erwartungen die Kursentwicklungen und müssen von Anlegern in die Entscheidungsfindung mit einbezogen werden (Schredelseker 2002, 398). Auch Haushaltsentscheidungen (z. B. Wahl des Verkehrsmittels zum Arbeitsplatz) können nicht isoliert getroffen werden, sondern müssen die Gegebenheiten mit einbeziehen und unter Berücksichtigung der Handlungen der anderen Beteiligten getroffen werden (Schlösser 2012, 310). Im Beauty-Contest-Game – als Simulation interaktiver Entscheidungssituationen in der Wirtschaftspraxis – kommen demnach durchaus Gedankengänge und Entscheidungen zum Tragen, wie sie auch im Wirtschaftsleben erforderlich sind. Insbesondere sind dies auch Fähigkeiten wie zum Beispiel das Denken in Kreislauf- und Ordnungszusammenhängen, die ökonomisch Gebildeten zugeschrieben werden (Kaminski/Eggert 2008).

Gemäß dieser Argumentation ist zu erwarten, dass insbesondere ökonomisch Gebildete die interaktive Situation analytisch richtig erfassen und spieltheoretische Gedankengänge entwickeln sowie dementsprechend eine niedrige Zahl wählen. Somit scheint das Beauty-Contest-Game, z. B. als Unterrichtseinstieg, zur Vertiefung oder zur Diskussion mikroökonomischer Themen, in der Lage zu sein, der Lehrkraft einen einfachen und schnellen Überblick über die Fähigkeit der Lernenden zu verschaffen, sich in ökonomischen Handlungssituationen zu orientieren und rationale Denkmuster anwenden zu können. Die These lautet:

Je niedriger die gewählte Zahl im Beauty-Contest-Game, desto höher der Stand der ökonomischen Bildung der Teilnehmenden.
--

### **3 Instrumente der Leistungsdiagnostik in der ökonomischen Bildung**

Ob das Beauty-Contest-Game tatsächlich als Indikator für das Ausmaß ökonomischer Bildung taugt, wird nun in den folgenden Kapiteln eruiert. Dazu werden in diesem Kapitel mit der klassischen Klausur und dem wirtschaftskundlichen Bildungstest nach dem Literacy-Konzept zwei assessmentorientierte Diagnostikverfahren kurz expliziert und deren Aussagekraft bzw. Eignung in Hinblick auf die Messung ökonomischer Bildung diskutiert. Alle drei Instrumente wurden in wirtschaftswissenschaftlichen Studiengängen einer Hochschule eingesetzt. Die Triangulation des Beauty-Contest-Game mit den beiden anderen Instrumenten dient dazu, Erkenntnisse über dessen Diagnostikgüte ableiten zu können.

#### **3.1 Wirtschaftswissenschaftliche Klausur**

Ob an Schulen oder an Hochschulen – schriftliche Prüfungen sind traditionelle Bestandteile eines Bildungsganges und überprüfen den Lernfortschritt. Wirtschaftswissenschaftliche Klausuren im Speziellen überprüfen dabei den Lernerfolg in Veranstaltungen, die einen breiten ökonomischen Wissensfundus sowie ein großes Instrumentarium zur Analyse ökonomischer Prozesse bereitstellen. Ob jedoch der Transfer von theoretischem Wissen in die Wirtschaftspraxis tatsächlich gelingt und ob die Studierenden die Inhalte einzelner Lehrveranstaltungen vernetzen sowie auf die Anwendbarkeit in konkreten Handlungssituationen überprüfen, wird durch einzelne Klausurergebnisse nicht abgebildet. In diesem Kontext sei die wirtschaftsdidaktische Differenz zwischen Wirtschaftswissenschaft und Wirtschaftspraxis nach Liening (1999, 309) erwähnt, die durch die Klausur nicht berücksichtigt wird, da eben keine Lebens- und/oder reale Handlungssituationen aufgegriffen werden. Damit wird deutlich, dass eine wirtschaftswissenschaftliche Klausur allenfalls als Indikator und nicht als Test ökonomischer Bildung anzusehen ist. Gleichwohl haben Klausuren pragmatische Vorzüge und eine umfassende Leistungsdiagnostik kann sie nicht entbehren (Heller 1984, 33). Aufgrund ihrer hohen Praxisrelevanz kommt der Klausur im vorliegenden Untersuchungsdesign daher eine wichtige Ergänzungsfunktion zu.

Stellvertretend für den Testansatz „wirtschaftswissenschaftliche Klausur“ kam in der Erprobung an einer Hochschule die Klausur „Mikroökonomie“ zum Einsatz. Da Klausuren als subjektive Verfahren zu den nicht messenden Methoden zählen und somit im Allgemeinen nicht objektiven Testgütekriterien genügen, ist auch die eingesetzte Klausur testtheoretisch kritisch zu betrachten, wenngleich den Gütekriterien durch das Vieraugenprinzip der Klausurkorrektur (Objektivität), der langjährigen Erfahrungen des Klausurenstellers

(Fehlerfreiheit, Reliabilität) und einer auf in den Modulhandbüchern dargelegten Veranstaltungslernzielen abgestimmten Klausur (Konstruktvalidität) Rechnung getragen wurde.

Die diese Inhalte und Lernziele prüfende Klausur, an der auch die Beauty-Contest-Game-Probanden der abschließenden Untersuchung rund um die verschiedenen Indikatoren teilnahmen, war eine Aufgabenklausur mit vier offenen Aufgaben, von denen die Studierenden in 90 Minuten drei Aufgaben bearbeiten sollten.

### 3.2 Wirtschaftskundlicher Bildungstest (WBT)

„Der Wirtschaftskundliche Bildungstest misst eine kontextualisierte kognitive Leistung, die als ökonomische Intelligenz bezeichnet werden kann.“ (Beck/Krumm 1998, 7) Mit diesen Worten stellen die Autoren des WBT den im Jahr 1998 in Deutschland veröffentlichten ersten wissenschaftlich erprobten Test zur Messung ökonomischer Grundfertigkeiten vor. Dabei handelt es sich um eine Adaption des aus den USA stammenden „Test of Economic Literacy“ (TEL) (Soper/Walstad 1987). Da Struktur, Aufbau und Inhalt des WBT somit entscheidend vom TEL und damit auch von amerikanischen Vorstellungen ökonomischer Bildung geprägt wurden, muss einer Thematisierung des WBT eine zusammenfassende Darstellung der Hintergründe des TEL vorangehen.

Charakteristisch für die ökonomische Bildung in den USA ist die Identifizierung von zunächst vier Kernthemenbereichen ökonomischer Bildung: ökonomische Grundlagen, Mikroökonomie, Makroökonomie und internationale Wirtschaftsbeziehungen. Diese vier Felder wurden dann in wirtschaftswissenschaftliche Grundkonzepte aufgeschlüsselt, anhand derer ökonomisches Wissen und ökonomische Fertigkeiten entwickelt werden sollen. Insgesamt wurden so 22 Basiskonzepte identifiziert (Saunders 2012, 16). Zudem beinhalten die im „Framework for Teaching Basic Economic Concepts“ gesetzten Standards neben Faktenwissen auch Analyse-, Problemlöse- und Beurteilungsfähigkeiten in konkreten ökonomischen Handlungssituationen (Hoidn/Kaminski 2006, 2). Dies erklärt, warum das Rahmenkonzept auch unter dem Schlagwort „Economic Literacy“ gefasst wird. Der Begriff „Literacy“ lässt sich gemäß UNESCO (2004, 13) wie folgt fassen:

*„Literacy is the ability to identify, understand, interpret, create, communicate and compute, using printed and written materials associated with varying contexts. Literacy involves a continuum of learning in enabling individuals to achieve their goals, to develop their knowledge and potential, and to participate fully in their community and wide society.“*



Mit Einführung des „Economic Literacy“-Konzepts wurden vorherige Tests grundlegend überholt und das Konzept als neuer Standard der TEL in den USA als aussagekräftigster Test zur Messung ökonomischer Bildung eingeführt. Dieser Multiple-Choice-Test orientiert sich dabei genau an den zuvor genannten vier Hauptthemenbereichen und 22 Basiskonzepten (Soper/Walstad 1987, 3). Der WBT hat als deutsche Adaption des TEL die Struktur und den Fragenkatalog des TEL genau übernommen und somit auch die dahinterstehenden Ideen des „Economic Literacy“-Konzepts nach Deutschland transferiert (Beck/Krumm 1998, 18). Wie der TEL ist auch der WBT ein Multiple-Choice-Messinstrument, das in zwei äquivalenten und sich zum Teil überschneidenden Testversionen „Form A“ und „Form B“ mit je 46 Multiple-Choice-Fragen vorliegt. Insgesamt stellt der WBT somit eine Aufgabenbank von 77 Multiple-Choice-Fragen zur Verfügung, die sich in Anlehnung an Blooms Klassifikation von Lernzielen (1954) den verschiedenen kognitiven Stufen „Wissen“, „Verstehen“, „Anwenden“, „Analyse“ und „Evaluation“ zuordnen lassen.

Der WBT ist im Gegensatz zur Klausur im Sinne der Testtheorie ein objektives Messinstrument und national wie international mit Abstand das meist verwendete Diagnostikverfahren in der ökonomischen Bildung. Allerdings weist auch der WBT in Hinblick auf die Messung ökonomischer Bildung Mängel auf. Der WBT zielt auf ökonomisches Grundlagenwissen ab und kann mit Erfolg absolviert werden, ohne dass dieses Wissen auf praktische Bedeutsamkeit anhand von eigenen Erfahrungen im Wirtschaftsalltag überprüft wurde. Somit spiegelt das Ergebnis im WBT eher wider, inwieweit mit fundiertem Grundlagenwissen die Voraussetzungen für ökonomische Bildung vorliegen. Es ist ein weiterer Indikator ökonomischer Bildung.

#### **4 Validierung des Beauty-Contest-Game**

Das Beauty-Contest-Game, die wirtschaftswissenschaftliche Klausur und ein nach den Vorgaben der WBT-Handanweisung zusammengestellter WBT-Fragen-Pool von 30 Fragen<sup>2</sup> wurden auf eine Probandengruppe an einer Hochschule angewandt. Zentrale Ergebnisse dieser Untersuchung werden nun zusammenfassend dargestellt. Insbesondere steht im Mittelpunkt, ob ein Zusammenhang zwischen den Ergebnissen in den einzelnen Testformen besteht. Gegenüber allen drei Testformen – so die Argumentation in den beiden vorangegangenen Kapiteln – können bildungstheoretisch bzw. lebenssituativ Bedenken geäußert

---

<sup>2</sup> In der Handanweisung wird explizit auf die Möglichkeit zum Einsatz einer gekürzten Fassung hingewiesen. Da sich der WBT auf eine mittlerweile überholte Version des TEL bezieht, wurden insbesondere Fragen mit fachlichen Ungenauigkeiten (etwa Fragen nach dem Lombardsatz) ausgeklammert.

werden. Allerdings spiegelt jeder Ansatz auf seine Art die Realität ökonomischer Bildung wider und kann zumindest als ein Indikator ökonomischer Bildung angesehen werden. Ob sich diese Gemeinsamkeit auch in der Untersuchung zeigt, soll insbesondere eine Korrelationsanalyse aufzeigen. So lassen sich Aussagen darüber treffen, ob die Zahlenwahl im Beauty-Contest-Game mit den Leistungen der Spezialwissen abfragenden Klausur oder mit dem inhaltlich eher allgemeineren WBT in Zusammenhang steht.

#### 4.1 Allgemeine Untersuchungsdeskription

Die Probandengruppe wurde zusammengestellt, indem Teilnehmende zufällig aus der Grundgesamtheit von über 500 Studierenden in mehreren Runden angeschrieben wurden, bis mehr als 100 ( $n = 115$ ) an einer Online-Befragung teilgenommen haben. Die Online-Befragung umfasste das Beauty-Contest-Game sowie ausgewählte Fragen des WBT. Über die Matrikelnummer wurde die Leistung in der kurz darauf stattfindenden Mikroökonomieklausur ermittelt. Details zur Zusammensetzung der Probandengruppe können Tabelle 1 entnommen werden. Die Analyse der Untersuchungsergebnisse wurde auf Basis einer Korrelationsanalyse durchgeführt.

#### 4.2 Zusammenhang WBT und Klausur

Vergleicht man zunächst die WBT-Ergebnisse (vgl. Abbildung 1) mit dem Gesamtergebnis der Klausur (vgl. Abbildung 2), so deutet sich bereits ein Zusammenhang an: Wenigen guten Ergebnissen stehen deutlich mehr im mittleren und schwachen Bereich gegenüber.

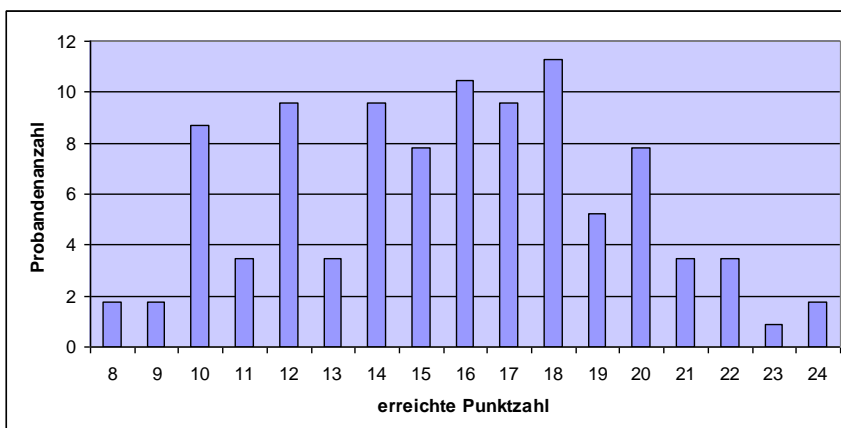


Abbildung 1: WBT-Punkteverteilung

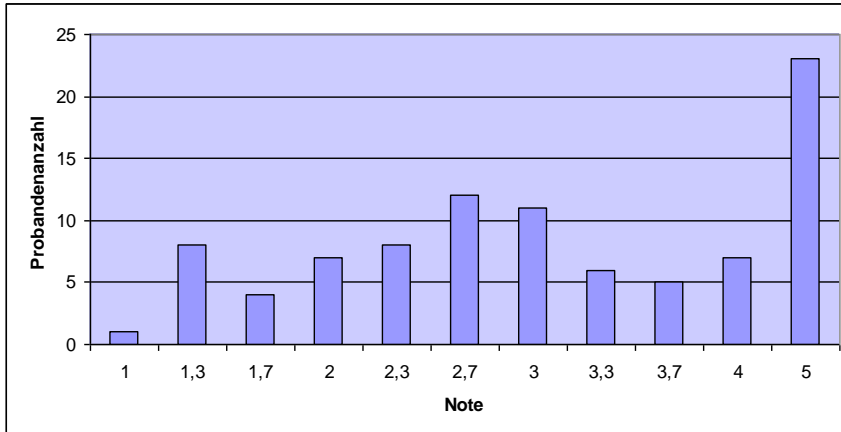


Abbildung 2: Notenverteilung in der Klausur „Mikroökonomie“

Im WBT beantworteten die Studierenden im Durchschnitt 52 % der Fragen richtig. Dieser Wert liegt deutlich unter dem Ergebnis der WBT-Normierungsstudie, in der Studierende der Wirtschaftswissenschaft von den in der Untersuchung zum Einsatz gekommenen 30 Fragen im Durchschnitt 75 % der Fragen richtig beantworteten. Das Ergebnis erscheint demnach und angesichts der Tatsache, dass die Fragen ökonomische Basiskonzepte abdeckten und lediglich Grundlagenwissen sowie gängige wirtschaftliche Grundfertigkeiten einforderten, für eine Studierendengruppe zu gering. Dass der Test tatsächlich gut differenzierte, zeigt die Punkteverteilung mit einer Bandbreite von mindestens acht bis höchstens 24 Punkten. Ähnlich wie im WBT ist das Leistungsniveau der Studierenden anhand der Klausur eher als schwach einzustufen (Durchschnittsnote: 3,23).

	WBT				Klausur			
	Anzahl	richtige Antworten (in %)	erreichte Punktzahl	Standardabweichung (SD)	Anzahl	Notenschnitt	SD	
alle Probanden	115	52 %	15,66	3,769	92	3,23	1,253	
Semesteranzahl	2	96	52 %	15,75	3,636	80	3,2	1,263
	4	13	55 %	16,46	4,465	8	3,03	1,134
	> 5	6	42 %	12,5	3,332	4	4,12	1,256
Geschlecht	m	61	54 %	16,33	3,876	51	3,15	1,253
	w	54	50 %	14,91	3,530	41	3,48	1,263

Tabelle 1: Übersicht der Probandenleistungen in WBT und Klausur

Nimmt man eine differenzielle Perspektive ein und teilt die Probandengruppe in spezifische Teilgruppen auf, so zeigen sich ebenfalls Gemeinsamkeiten, wenngleich bezogen auf die Klausur keine statistische Signifikanz gegeben ist. Wie Tabelle 1 entnommen werden kann, erzielen die Studierenden des 4. Semesters in beiden Testverfahren bessere Leistungen als

die des 2. Semesters, während Studierende des 5.-12. Semesters deutlich abfallen – im WBT gilt dabei eine Signifikanz von 0,05. Zudem werden auch bezogen auf das Geschlecht im WBT und in der Klausur gleiche Befunde hervorgebracht: Männliche Probanden erzielen bessere Leistungen als weibliche. Der geschlechterspezifische Unterschied im WBT ist statistisch signifikant zum Niveau von 0,05.

Entscheidender für die Beurteilung, ob die Leistungen in WBT und Hochschulklausur im Zusammenhang stehen, ist eine Korrelationsanalyse, in der 92 Wertepaare in der Korrelationsanalyse verwertet werden konnten. 23 Studierende, die die Fragen des WBT beantworteten, nahmen nicht an der Klausur teil. Die Analyse der Wertepaare ( $x$ ,  $y$ ) der Probanden mit „ $x$  = erreichte Punktzahl im WBT“ und „ $y$  = Note in der Mikroökonomieklausur“ ergibt einen Korrelationskoeffizienten von  $r = -0,532$  und weist einen (negativen) linearen Zusammenhang der Ergebnisse der beiden Testansätze nach. Gemäß der Klassifizierung von Bortz und Döring (2006, 606) ist er mit  $|r| \geq 0,5$  als stark einzustufen. Dabei ist die Korrelation auf einem Niveau von 0,01 statistisch signifikant. Zwar sagt eine Korrelation nichts über die Kausalität des Zusammenhangs aus; unter Rückgriff auf die theoretischen Ausführungen, welche betonten, dass ökonomisch Gebildete in beiden Testansätzen potenziell gute Leistungen erbringen können, lässt sich vermuten, dass eine dieser gemeinsamen Ursachen im Stand der ökonomischen Bildung liegt. Es bleibt festzuhalten: Leistungen in WBT und Hochschulklausur stehen in signifikant starkem Zusammenhang. An dieser Stelle sei ein damit einhergehender Befund von Wuttke und Beck (2002, 8) erwähnt, wonach die Ergebnisse im WBT mit den Diplomnoten im wirtschaftswissenschaftlichen Studium, welche letztendlich eine Aggregation von einzelnen Klausurleistungen darstellen, in einem deutlichen Zusammenhang stehen.

### **4.3 Zusammenhang des Beauty-Contest-Game mit WBT und Klausur**

Erste Anhaltspunkte für einen Zusammenhang der drei Instrumente können gewonnen werden, indem das Gesamtergebnis und die Leistungen spezifischer Teilgruppen mit den Befunden aus WBT und der Klausur verglichen werden. Der Durchschnitt aller gewählten Zahlen der Probandengruppe an der Hochschule ist die Zahl ZA: 32,91 (SD: 18,772) und dementsprechend lautet die Gewinnzahl  $2/3 \times 32,91 = 21,94$ , die höher ist als die in 2.1.3 vorgestellten und von Fachzeitschriften durchgeführten Untersuchungen.

Deutet man hohe Zahlen im Sinne der Hypothese als Anzeichen von Defiziten in grundlegenden ökonomischen Denkmustern, so ist auch dieses Ergebnis wie beim WBT und der Hochschulklausur als eher schwach einzustufen. Auch die Verteilung der gewählten Zahlen (vgl. Abbildung 3) ähnelt der Punkte- bzw. Notenverteilung in Abbildung 1 und 2:

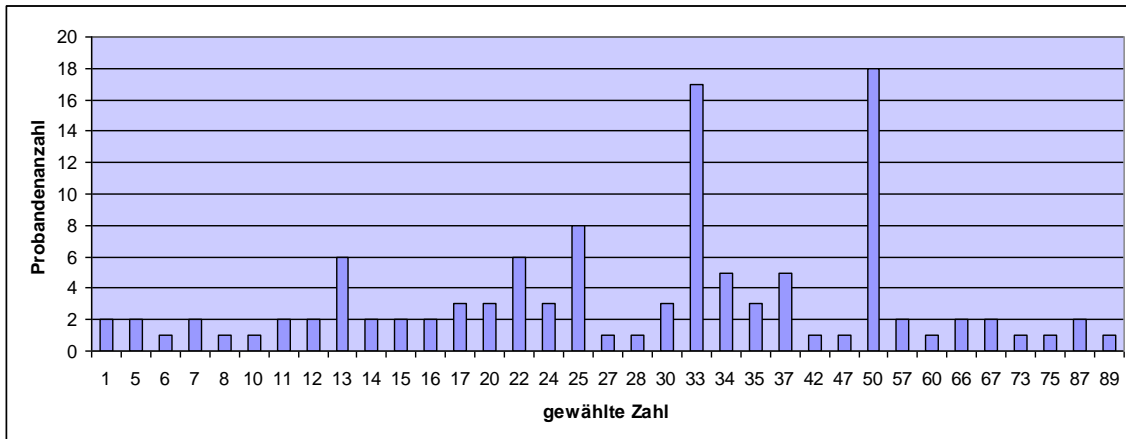


Abbildung 3: Verteilung der im Beauty-Contest-Game gewählten Zahlen

Wenigen guten Leistungen (d. h. eine niedrige Zahl) stehen viele Zahlen im mittleren und hohen Bereich gegenüber. Auffällig sind die Spitzen bei den Zahlen 50 und 33, die auch Nagel und Bosch-Domènech in den in 2.1.3 vorgestellten Experimenten beobachteten und darauf zurückführen, dass entweder der Durchschnitt aller Zahlen (= 50) gewählt wurde oder ausgehend von dieser Zahl lediglich ein rekursiver Gedankengang ( $2/3 \times 50 = 33,3\bar{3}$ ) durchgeführt wurde (Nagel/Bosch-Domènech 1999, 10).

Erneut zeigen sich geschlechtsspezifische (jedoch nicht signifikant starke) Unterschiede zwischen männlichen Probanden (ZA: 31,74/ SD: 17,656) und weiblichen (ZA: 34,24/ SD: 20,041). Lediglich bei der Differenzierung nach Semesteranzahl stimmen die Befunde des Beauty-Contest-Game nicht mit vorangegangenen Resultaten überein: Studierende des 2. Semesters (ZA: 31,89/ SD: 16,024) haben sich im Durchschnitt geschickter verhalten als die des 4. Semesters (ZA: 34,31/ SD: 32,04). Der Unterschied ist statistisch jedoch nicht signifikant.

Die in Hinblick auf die Einsatzfähigkeit des spieltheoretischen Experimentes in der Lerndiagnostik durchgeführte Korrelationsanalyse der Wertepaare (x, z) mit „x = erreichte Punktzahl im WBT“ und „z = Zahlenwahl im Beauty-Contest-Game“ ergibt einen Korrelationskoeffizienten von  $r = -0,289$  und weist einen zum Niveau von 0,01 signifikanten (negativen) linearen

Zusammenhang der Zahlenwahl mit den Ergebnissen des WBT nach, der gemäß der Klassifizierung von Bortz und Döring (2006, 606) als „schwach“ und nur knapp unter „mittel“ einzuordnen ist. Dies bestärkt die bereits formulierte Vermutung bezüglich der Fehleranfälligkeit des Indikators Beauty-Contest-Game. Gleichwohl ist aber zu unterstreichen, dass ein signifikanter, nicht trivialer Zusammenhang besteht.

Die Korrelationsanalyse der Wertepaare  $(y, z)$  mit „ $y$  = Note in der Mikroökonomieklausur“ und „ $z$  = Zahlenwahl im Beauty-Contest-Game“ liefert eine zum Niveau von 0,01 signifikante Korrelation mit  $r = 0,425$ , d. h. eine mittlere Effektstärke (Bortz/Döring 2006, 606). Da sowohl Mikroökonomieklausur als auch Beauty-Contest-Game einen signifikanten Zusammenhang zum WBT aufwiesen, konnte im Sinne einer triangularen Messung erwartet werden, dass die beiden Indikatoren Beauty-Contest-Game und Mikroökonomieklausur ebenfalls im Zusammenhang stehen. Der nach dem Williams-Test (Meng et al. 1992) zum Niveau von  $< 0,001$  signifikant stärkere Zusammenhang zwischen Beauty-Contest-Game und Klausur ist vermutlich darauf zurückzuführen, dass die im Beauty-Contest-Game zum Tragen kommenden spieltheoretischen Denkmuster mit der Mikroökonomieklausur eine größere Schnittmenge aufweisen als die von Beauty-Contest-Game und WBT. Einen Überblick über alle in diesem Kapitel ermittelten Korrelationskoeffizienten gibt die folgende Tabelle. Alle Korrelationen sind dabei zum Niveau von 0,01 statistisch signifikant.

<i>Indikator</i>	<b>WBT</b>	<b>Klausur</b>	<b>Beauty-Contest-Game</b>
<b>WBT</b>	1	-0,532	-0,289
<b>Klausur</b>	-0,532	1	0,425
<b>Beauty-Contest-Game</b>	-0,289	0,425	1

Tabelle 2: Übersicht ermittelter Korrelationskoeffizienten

## 5 Limitationen und Diskussion

Der vorliegende Beitrag zeigt die Möglichkeit einer Lerndiagnostik im Wirtschaftsunterricht auf, die nicht nur auf leistungsorientierte Verfahren, sondern auch auf prozessorientierte Diagnostikinstrumente der experimentellen Ökonomie zurückgreift. Dazu wurde mit dem Beauty-Contest-Game ein spieltheoretisches Unterrichtsexperiment adaptiert und argumentativ dargelegt, wie dieses als Unterrichtseinstieg, aber auch als vertiefendes, prozessbegleitendes Diagnoseinstrument im Sinne eines Indikators zur Fähigkeit rationalen Denkens in ökonomischen interaktiven Entscheidungssituationen im Wirtschaftsunterricht zum Einsatz kommen kann.

Gleichwohl wurde darauf hingewiesen, dass die Einsatzmöglichkeiten eines solchen spieltheoretischen Experimentes gewissen Grenzen unterliegen. Das Konstrukt „Ökonomische Bildung“ kann in all seinen Facetten selbstverständlich nicht durch ein isoliert durchgeführtes Experiment erfasst werden. Die Verwendung spieltheoretischer Aufgaben kann jedoch als zusätzliche Informationsquelle bzw. Indikator ökonomischer Bildung betrachtet werden. Ebenso wie der lerndiagnostische Einsatz des Beauty-Contest-Game Grenzen unterliegt, ist auch die im Rahmen dieses Beitrags durchgeführte Validierungsuntersuchung gewissen Limitationen ausgesetzt. Bosshardt und Kennedy (2012) definieren einen „Goldstandard“ für empirische Forschungsvorhaben in der ökonomischen Bildung. Dort wird aufgezeigt, dass insbesondere die Zusammensetzung der Probandengruppe genauer zu betrachten ist. In der vorliegenden Untersuchung wurden die Probanden zwar zufällig aus der Grundgesamtheit der Studierenden gezogen, jedoch war die Durchführung auf die Teilnahmebereitschaft der Studierenden angewiesen. So ist etwa zu erwarten, dass tendenziell engagierte Studierende am freiwilligen Teil der Untersuchung teilnahmen (Beauty-Contest-Game und WBT). Hinsichtlich der eingesetzten Diagnoseinstrumente und dessen Vergleichbarkeit ist ferner anzumerken, dass den Probanden bei der Bearbeitung der angesetzten Klausur, anders als bei der Durchführung des WBT und des Beauty-Contest-Game, eine persönliche Leistungsmotivation zugrunde lag, sodass die Klausur möglicherweise mit besonderer Konzentration bearbeitet wurde. Auch die Tatsachen, dass die Untersuchung sich auf die Studierendengruppe einer einzelnen Hochschule bezieht und eine bestimmte Beispielklausur dieser Hochschule zum Einsatz kam, zeigen Grenzen bei der Interpretation des Untersuchungsergebnisses auf. Zudem wurde von signifikanten Zusammenhängen zwischen den eingesetzten Testansätzen berichtet, aufgrund derer das Beauty-Contest-Game als zusätzliche Informationsquelle im inter- und intragruppenspezifischen Diagnostikprozess herangezogen wurde, der so durch interaktive Elemente erweitert werden kann. Gleichwohl waren diese Zusammenhänge jedoch nie so stark ausgeprägt, dass das betrachtete Beauty-Contest-Game als isolierter valider Test gelten kann. Dazu wären Korrelationskoeffizienten betragsmäßig nahe 1 nötig gewesen.

Festzuhalten bleibt also, dass unter Beachtung dieser Limitationen ein signifikanter, nicht trivialer (d. h. mittlerer bis schwacher) Zusammenhang zwischen dem Spielverhalten im Beauty-Contest-Game und den Leistungen in zwei weiteren wirtschaftsdidaktischen Diagnostikinstrumenten festgestellt wurde. Der Beitrag liefert damit einen Impuls, spieltheoretische und damit interaktive Unterrichtsexperimente als zusätzliche Informationsquellen für Indikatoren in der Lerndiagnostik ökonomischer Bildung verstärkt zu berücksichtigen.

## Danksagung

Unsere besondere Anerkennung gilt den am Review-Prozess Beteiligten, die mit konstruktiver Kritik wesentlich zur Stringenz des Aufsatzes beigetragen haben. Ferner wäre ohne das Engagement der ungenannt bleibenden Mikroökonomie-Lehrenden und -Studierenden die Umsetzung nicht möglich gewesen. Für die Unterstützung herzlichen Dank.

## Literaturverzeichnis

- Beck, K., Krumm, V. (1998): Wirtschaftskundlicher Bildungs-Test. Handanweisung, Göttingen.
- Berninghaus, S., Ehrhart, K., Güth, W. (2010): Strategische Spiele. Eine Einführung in die Spieltheorie, Heidelberg, 3. Aufl.
- Bloom, B. S. (1954): Taxonomy of educational objectives. Handbook I: cognitive domain, New York.
- Bortz, J., Döring, N. (2006): Forschungsmethoden und Evaluation, Heidelberg, 4. Aufl.
- Bosch-Domènech, A., Nagel, R. (1997a): El juego de adivinar el número X: una explicación y la proclamación del vencedor. In: *Expansión*. o. Jg., 6/1997, 42f.
- Bosch-Domènech, A., Nagel, R. (1997b): Guess the Number: Comparing the F's and Expansion's Results. In: *Financial Times*. o. Jg., 30. Juni 1997, 14.
- Bosshardt, W., Kennedy, P. (2012): Data resources and econometric techniques. In: Hoyt, G./McGoldrick, K. (Hg.): *International Handbook on Teaching and Learning Economics*, Cheltenham, UK, 373-383.
- Heller, K. A. (1984): *Leistungsdiagnostik*, Bern.
- Hoidn, S., Kaminski, H. (2006): Ökonomische Bildung in den USA. In: *Trends in Bildung international*, 12, 1-9.
- Keynes, J. M. (1936): *Allgemeine Theorie der Beschäftigung, des Zinses und des Geldes*, Berlin.
- Kramer, K. (2002): *Lernkompetenzen beobachten, beurteilen und fördern: Kriterienkatalog und Fragebogen zur Einschätzung der Methoden-, Sozial-, Selbst- und Sachkompetenz bei Schülerinnen und Schülern unter der besonderen Berücksichtigung von grundlegenden Voraussetzungen für Lebenslanges Lernen*, Kiel.
- Liening, A. (1999): *Komplexe Systeme zwischen Ordnung und Chaos – Neuere Entwicklungen in der Theorie nicht-linearer dynamischer Systeme und die Bedeutung für die Wirtschaftswissenschaft und ihre Didaktik*. Zugel. Habil.: WWU Münster, Münster.
- Meng, X., Rosenthal, R., Rubin, D. (1992): Comparing correlated correlation coefficients. In: *Psychological Bulletin*, Vol. 111 (1), 172-175.
- Mittelstädt, E. (2011): *Ökonomische Schulentwicklung — Wissensbilanzierung zum angemessenen Umgang mit Komplexität*. Zugel. Diss.: TU Dortmund, Frankfurt/Main.



- Nagel, R. (1999): A Survey on Experimental Beauty Contest Games: Bounded Rationality and Learning. In: Budescu, V./Erev, I./Zwick, R. (Hg.): Games and Human Behaviour – Essays in Honor of Amnon Rapoport, Mahwah 1999, 105ff.
- Nagel, R., Bosch-Domènech, A. (1999): One, Two, Three, Infinity: Newspaper & Lab Beauty Contest Experiments 1999, 10 ff. AVL <http://www.econ.upf.edu> (2009-09-04).
- Retzmann, T. (2008): Ökonomische Bildung und Spieltheorie. In: Loerwald, D., Wiesweg, M., Zoerner, A. (Hrsg.): Ökonomik und Gesellschaft. Festschrift für Gerd-Jan Krol, Wiesbaden.
- Ross, L., Greene, D., House, P. (1977): The „False Consensus Effect“: An Egocentric Bias in Social Perception and Attribution Processes. In: Journal of Experimental Social Psychology, 13, 279-301.
- Saunders, P. (2012): A history of economic education. In: Hoyt, G., McGoldrick, K. (Hg.): International Handbook on Teaching and Learning Economics, Cheltenham, UK, 3-36.
- Schlösser, H.-J. (2012): Homo oeconomicus. In: May, H., Wiepcke, C. (Hg.): Lexikon der ökonomischen Bildung, 8. Aufl., München, 310-312.
- Schrader, F.-W. (1989): Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts, Frankfurt.
- Schredelseker, K. (2002): Grundlagen der Finanzwirtschaft, Oldenburg.
- Selten, R., Nagel, R. (1998): Das Zahlenwahlspiel – Ergebnisse und Hintergründe. In: Spektrum der Wissenschaft, o. Jg. (2), 1998, 16-22.
- Soper, J. C., Walstad, W. B. (1987): Test Of Economic Literacy – Second Edition, Examiner's Manual, New York.
- Terhart, E. (2002): Standards für die Lehrerbildung – Eine Expertise für die Kultusministerkonferenz. AVL <http://www.sowi-online.de> (2012-07-31).
- UNESCO (2004): The Plurality of Literacy and its implications for Policies and Programs. Position Paper. AVL <http://unesdoc.unesco.org> (2009-09-04).
- Wuttke, E., Beck, K. (2002): Eingangsbedingungen von Studienanfängern – Die prognostische Validität wirtschaftskundlichen Wissens für das Vordiplom bei Studierenden der Wirtschaftswissenschaften. In: Arbeitspapiere WP, Nr. 41, Mainz.