

Evaluierung von Lernmanagement- Systemen (LMS): Theorie – Durchführung – Ergebnisse

Peter Baumgartner, Hartmut Häfele, Kornelia Maier-Häfele

Baumgartner, Peter; Häfele, Hartmut; Maier-Häfele, Kornelia (2002): „Evaluierung von Lernmanagement-Systemen: Theorie - Durchführung - Ergebnisse“. In: Hohenstein, Andreas; Wilbers, Karl (Hrsg.) Handbuch E-Learning. Köln: Fachverlag Deutscher Wirtschaftsdienst, Kap.5.4, 1. Erg.-Lfg.

Im Auftrag des österreichischen Bundesministeriums für Bildung, Wissenschaft und Kultur (bm:bwk) haben wir einen weltweiten Vergleich von Lernplattformen durchgeführt. Der nachfolgende Artikel beschreibt zuerst das allgemeine methodische Konzept, dann den konkreten Ablauf des Verfahrens und fasst im letzten Abschnitt die Ergebnisse dieses Evaluierungsprozesses zusammen.

Das Evaluationsverfahren (Theorie und Konzept)

Als Methode für das Evaluierungsverfahren wurde die sogenannte Qualitative Gewichtung und Summierung (QGS) gewählt. Die Begründung dieser Methode wurde bereits am Beispiel zweier Preisausschreiben - EASA [W1] und MeDiDa-Prix [W2] – sowohl theoretisch als auch in ihrer praktischen Handhabung ausführlich beschrieben (Baumgartner/Payr 1997, Baumgartner/Frank 2000). Da das Verständnis der Methode für die Diskussion der Ergebnisse des von uns durchgeführten weltweiten Vergleichs von Lernplattformen entscheidend ist, fassen wir das theoretische Konzept hier nochmals zusammen:

Es ist Standard, dass bei vergleichenden (Produkt-)Untersuchungen Kriterienkataloge eingesetzt werden. In Form von Prüf- oder Checklisten soll eine Bewertung der Qualität der zu evaluierenden Produkte erfolgen. Für die große Verbreitung dieses Verfahrens sind mehrere Vorteile verantwortlich:

- ❶ *Kostengünstig:* Es genügt eine fachkundige Person, ein Testaccount bzw. eine Programmkopie und eine Hardwareausstattung, auf der die Software mit all ihren Attributen lauffähig ist.
- ❷ *Schnell:* Durch die Trennung des Medienproduktes von seinem realen Einsatzgebiet (dem Lehrgeschehen) kann die Bewertung rasch erfolgen.
- ❸ *Methodisch sauber:* Durch die schrittweise, immer gleiche Abarbeitung umfangreicher Kriterienlisten erscheint das Verfahren objektiv und methodisch sauber.

So wichtig und verdienstvoll die Bewertung mittels Kriterienkatalogen auch sein mag, so weist diese Vorgangsweise auch einige gravierende methodische Mängel auf (vgl. dazu auch: Biermann 1994), von dem hier das Gewichtungsproblem herausgegriffen werden soll.

Obwohl über die Items eines Kriterienkatalogs relativ schnell Einigkeit geschaffen werden kann, ist die Frage, wie nun die einzelnen Faktoren gewichtet werden sollen, meist sehr umstritten. Gerade die Gliederung und Gewichtung der einzelnen Kriterien entscheidet aber über die vergleichende Bewertung und Auswahlentscheidung.

Es gibt zwei Verfahren zur relativen Bewertung einzelner Items eines Kriterienkatalogs - *numerische* und *qualitative* Gewichtungsprozeduren:

Die Numerische Gewichtung und Summierung (NGS)

Numerisches Gewichten und Summieren (NGS) stellt derzeit das dominante Modell für komplexe Produktevaluationen dar und wird insbesondere im Zusammenhang mit der Bewertung von Lernsoftware in Form von Check- oder Prüflisten angewendet (vgl. Baumgartner 1997; Biermann 1994; Doll 1987; Fricke 1995; Thomé 1988). NGS kommt in verschiedenen Formen vor und kann sowohl beschreibend (normativ, präskriptiv) oder auch bewertend eingesetzt werden. Die allgemeine Form ist die Multi-Attribute Utility Analysis (Scriven 1991a: 380f) Die Prozedur dabei ist folgende:

- ❶ Zuerst werden die einzelnen Dimensionen in ihrer relativen Wertigkeit (z.B. anhand einer 1 bis 3-, 1 bis 5- oder 1 bis 10-Skala) eingeschätzt bzw. gewichtet.
- ❷ Anschließend wird die Leistung des Evaluanden (= die zu evaluierende Sache, z.B. Lernsoftware) nach den einzelnen Dimensionen eingeschätzt (rating).
- ❸ Das Produkt von Leistungsbewertung und Gewicht (Leistungspunkte x Gewichtung) wird berechnet und für jeden einzelnen Evaluanden summiert.
- ❹ Es ergibt sich für jeden Evaluanden eine einzige Zahl, die den relativen Rang des jeweiligen Evaluanden bestimmt. Sieger ist der Evaluand mit der größten Punktezahl.

Das NGS-Verfahren ist infolge einer Reihe von Vorteilen (leicht verständlich, einfach durchzuführen, aufschlussreich, ergibt immer eine Rangordnung) sehr beliebt. Obwohl es einen ersten Aufschluss bzw. Einblick bietet und daher im Rahmen einer weiterführenden Evaluation durchaus brauchbar ist, hat dieses Verfahren jedoch schwerwiegende methodische Mängel, so dass die alleinige Verwendung dieses Verfahrens ungeeignet ist:

Minimum-Problem: Ein Set von Gewichten löst nicht das Problem, dass einige Dimensionen (Merkmale) erst dann eine sinnvolle Funktion des Evaluanden darstellen, wenn ein bestimmtes Mindestmaß überschritten ist. In einer abschließenden Summierung zu einer einzigen Zahl gehen diese inhaltlichen Minimalanforderungen jedoch verloren. Diese Schwierigkeit lässt sich jedoch durch eine Erweiterung des NGS-Verfahrens beheben (NGS-Modell mit Minima). Dabei wird jedes der Minima-Kriterien zuerst geprüft, bevor die Analyse fortgesetzt wird. Nur eine

Leistung über dem erforderlichen Minimum wird gewichtet. Evaluanden, bei denen einzelne Kriterien dieses notwendige Minimum nicht erreichen, scheiden aus (K.O.-Kriterien).

Problem der isolierten Kriterien: Ein weiteres (lösbares) Problem der NGS-Methode besteht darin, dass die einzelnen Bewertungskomponenten der Evaluanden oft nicht unabhängig voneinander zu betrachten sind, weil sie zueinander in Beziehung stehen und miteinander interagieren. Diese Schwierigkeit könnte durch eine Neubestimmung bzw. neue Definition der Kriterien gelöst werden. Allerdings ist dies nicht immer einfach, erfordert große Geschicklichkeit und Kenntnisse und stellt fast immer nur eine ad hoc-Lösung dar, die nicht verallgemeinert werden kann.

Annahme einer linearen Skala: Die NGS-Methode nimmt eine lineare Skala der Nützlichkeit (Vergabe von Punkten und Summierung) an, was jedoch falsch ist. Die verschiedenen Komponenten des Evaluanden lassen sich nicht über eine einzige Skala bewerten. Ähnlich wie beim erweiterten NGS-Modell (mit Minima) lässt sich auch hier durch ein sequentiell durchgeführtes Ausscheidungsverfahren Abhilfe schaffen: Es wird die Liste der Merkmale nicht zufällig (alphabetisch oder nach einer anderen inhaltlich irrelevanten Reihenfolge) durchgearbeitet, sondern zuerst werden die absoluten Notwendigkeiten festgestellt und dann so viele Kandidaten wie möglich eliminiert. Allerdings bleibt die grundsätzlich falsche Annahme einer Linearität der Punkteabstände im weiteren Verfahren bestehen. Multiplikation und Summenbildung sind nur bei Intervall- oder Ratio-Skalen zulässige Operationen, während es sich hier um eine Ordinalskala handelt, die nur eine Reihung der einzelnen Merkmale erlauben würde.

Problem des Detaillierungsgrades: Die entscheidende Kritik bzw. das (unlösbare) Hauptproblem des NGS-Verfahrens besteht jedoch darin, dass die Anzahl der Kriterien nicht voraussehbar ist. Sie kann von etwa einem Dutzend bis zu einigen hundert Kriterien reichen. Damit werden aber entweder wichtige Dimensionen durch eine Vielzahl von Trivialitäten überschwemmt oder aber weniger wichtige Faktoren wirken sich auf das Gesamtergebnis zu stark aus. Das Festlegen einer fixen Punkteanzahl, die nicht überschritten werden darf, reduziert zwar das Problem, kann es aber nicht gänzlich lösen. Was sind die relevanten Kriterien (Wie viele? Wie detailliert?) und welche Gewichtung kommt ihnen jeweils zu?

Besonders fatal beim NGS-Verfahren ist es, dass diese Gewichtungsprozedur keine Spuren hinterlässt. Da sich als Ergebnis bloß eine einzige Zahl pro Evaluand ergibt, sind nachträglich keine inhaltlichen Fehlerkorrekturen mehr möglich.

Die Qualitative Gewichtung und Summierung (QGS)

Obwohl mit dem NGS-Verfahren zwar viele Evaluanden in einem ersten Durchgang provisorisch miteinander verglichen werden können und es als erster grober Filter durchaus brauchbar ist, ist es letztlich notwendig, einen paarweisen Vergleich mit qualitativen Bewertungsverfahren durchzuführen (Scriven 1991a: 293ff):

In einem ersten Schritt werden für die einzelnen Dimensionen nur fünf Gewichte vergeben. Es empfiehlt sich dafür nicht-ziffernartige Symbole zu verwenden, damit gleich zu Beginn eine

Verwendung als Intervall- oder Ratioskala ausgeschlossen wird. Bewährt hat sich folgende Einteilung: Essential (E) / Very Valuable (*) / Valuable (#) / Marginally Valuable (+) / Zero (0). Damit werden nicht nur die Gewichte der einzelnen Merkmale festgelegt, sondern auch festgelegt, welche Eigenschaften Minimalerfordernisse (Essentials) darstellen.

Alle 0-Dimensionen können nun gestrichen werden. Sie sind als völlig unbedeutend gewichtet worden und daher für die weitere Bewertung irrelevant. Damit wird unnötiger Analyseaufwand vermieden.

Es wird nun überprüft, ob alle Evaluanden die Minimalerfordernisse (Kriterien, die mit „E“ gewichtet wurden) auch tatsächlich erfüllen. Falls nicht, werden sie aus der weiteren Analyse ausgeschieden. Dadurch wird der weitere Arbeitsaufwand beträchtlich reduziert. Allerdings ist dafür Sorge zu tragen, dass es sich dabei um ein alles-oder-nichts-Attribut handelt (z.B. die Lernplattform ist in Form einer nationalen Rahmenlizenz einsetzbar ja/nein). Andernfalls muss das (Anspruchs-)Niveau, das unbedingt erforderlich ist, genau festgelegt werden und geprüft werden, ob der betreffende Evaluand dieses Anspruchsniveau erreicht oder nicht. So kann z.B. ein Evaluand ein bestimmtes Minimum nicht nur erreichen (= E), sondern überschreiten und wird dann z.B. mit einem + versehen.

Die verbleibenden Evaluanden weisen jetzt nur mehr Unterschiede zwischen * und + auf und werden nun im Rahmen von 0 bis zur maximalen Gewichtung des jeweiligen Kriteriums bewertet. D.h. ein #-Kriterium kann keinen höheren Wert als # erhalten (also nur 0,+,#). Es besteht jedoch keine unbedingte Notwendigkeit, Bereiche für jeden Nützlichkeitslevel zu spezifizieren, d.h. einige Nutzensbereiche können auch übersprungen werden. So ist es z.B. möglich, dass eine Dimension nur + und * kennt, der Bereich mit # wird nicht einbezogen. Zu beachten ist außerdem, dass es Fälle gibt, in denen es keine monotonen Zweckmäßigkeitssbereiche gibt, d.h. das Überschreiten eines bestimmten Niveaus wird wiederum zu einer Schwäche (z.B. es gibt derart viele unterschiedliche und unverbundene asynchrone Kommunikationstools, dass die Bedienung – wann welches Tool verwendet werden soll – sehr kompliziert wird). Falls es Unsicherheiten über die Zuverlässigkeit (Reliabilität) der Einschätzung eines Leistungsmerkmals gibt, kann das Symbol eingeklammert werden. Damit kann der Evaluator/die Evaluatorin die Sicherheit der jeweiligen Beurteilung ausdrücken und das betreffende Kriterium wird damit für eine spätere – eventuell notwendig gewordene – genauere Untersuchung markiert.

Nach den bisherigen Verfahrensschritten entsteht nun eine Rangordnung (ranking), die anschließend auch mit einer integrierenden Schlussbewertung (grading) versehen werden kann (z.B. „kommt in die TOP 15“ etc.). Diese ließe sich z.B. durch Festlegung einer Minimumanzahl von * oder * und # oder auf einer individuellen Fallbasis durchführen, nachdem alle Evaluanden bereits bewertet wurden.

Nun werden die Ergebnisse der Leistungsbewertung integriert, indem jede Kategorie mit der gleichen Wertigkeit summiert wird, d.h. man erhält drei Gesamtwerte für jeden Evaluanden (= Summe der *, Summe der # und Summe der +, mit oder ohne Akzent, mit und ohne Klammer)

Im nächsten Schritt werden nun jene Eigenschaften, die alle Evaluanden gleichermaßen aufweisen (z.B. wenn alle Evaluanden ein bestimmtes Kriterium mit + erfüllt haben), ausgeschlossen. Damit wird der weitere Vergleich auf einer Fall-zu-Fall-Basis vereinfacht.

Abschließend kann nun geprüft werden, ob bereits eine eindeutige Rangordnung möglich ist. Eindeutig heißt, dass z.B. ein Evaluand mit 3*, 4# und 2+ auf jeden Fall besser ist als einer mit 2*, 4# und 3+. Hat jedoch der zweite Evaluand z.B. 2*,7#, so ist keine eindeutige Entscheidung möglich und die beiden Kandidaten müssen in einem paarweisen Vergleich genauer untersucht werden.

Wenn nun nicht bereits entscheidbare Verhältnisse vorliegen, so kann eine neuerliche Gewichtung im Lichte der vergleichenden Bewertung konkreter Einzelfälle hilfreich sein.

Neben seiner relativen Komplexität hat daher das QGS-Verfahren den Nachteil, dass es keinen definitiven Entscheidungsalgorithmus hat. Manchmal muss es als iterative Prozedur mehrfach durchlaufen werden und im Lichte der bisherigen Analyse müssen einzelne Bewertungen nochmals durchgeführt werden. Das Verfahren wechselt damit ständig zwischen holistischer und analytischer Betrachtungsweise, ergibt aber immer sinnvolle und vor allem nachvollziehbare und überprüfbare Ergebnisse.

Das Evaluationsfahren (praktische Durchführung)

Das im vorliegenden Fall vorgestellte Evaluierungsverfahren gliedert sich in zwei Abschnitte:

- Weltweiter Vergleich aller relevanten Lernplattformen und Erstellung einer Liste der Top 15: das Ergebnis ist eine qualitative Beschreibung dieser 15 Produkte.
- Detaillierter Vergleich der Produkte aus der Top-Liste. Installation von 5 Produkten und intensiver Usability-Test.

Dieser Artikel beschreibt den ersten Abschnitt (Mai 2001 bis Dezember 2001), der zweite Abschnitt des Evaluierungsverfahren (Jänner 2002 – Oktober 2002) ist zum gegenwärtigen Zeitpunkt (Anfang Februar) gerade angelaufen.

Der erste Teil des Verfahrens selbst wiederum lässt sich in drei verschiedene Phasen einteilen:

- Erarbeiten der Evaluierungskriterien inklusive Erstellung der Mindestkriterien (K.O.-Kriterien, Essentials) und qualitative Gewichtung der Kriterien
- Sichtung und Dokumentation der am Markt befindlichen Lernplattformen
- Erstellen der Shortlist (Top 15)

❶ Die Ausarbeitung der Evaluierungskriterien

Unser erster Schritt im Verfahren war es, Websites aufzubauen, über die wir vom Fortgang unserer Untersuchung kontinuierlich berichten. Es gibt dabei sowohl eine Website zum Gesamtprojekt und unseren detaillierten Erfahrungen (Projekttagbuch [W3]) als auch eine Community-Website [W4], auf der wir die Untersuchungsergebnisse dokumentieren.

Im Zusammenhang mit dem zweiten Schritt, dem Aufbau einer ExpertInnengruppe, war dies für einen geordneten und ungestörten Verlauf (relativ wenige Interventionen seitens unterschiedlicher Stakeholder wie z.B. Herstellerfirmen von Lernplattformen, Verantwortliche für eLearning in Organisationen, die bereits auf eine bestimmte Lernplattform „eingeschworen“ waren) des Evaluierungsverfahrens sehr wichtig.

Unser zweiter vorbereitender Schritt im Verfahren war es – gemeinsam mit dem Auftraggeber – eine Expertengruppe aufzubauen, über die sowohl die Art der Kriterien, als auch ihre relative Gewichtung zueinander abgestimmt wurden. Nachträglich gesehen war dies ein zentraler Punkt nicht nur für die Akzeptanz der Ergebnisse sondern auch für das Gelingen des – insgesamt sehr heiklen – Verfahrens.

Die Expertinnengruppe setzt sich aus VertreterInnen des Auftraggebers und von in Österreich relevanten Bildungsorganisationen verschiedener Bereiche (Schulen, Fachhochschulen, Universitäten und Weiterbildung) zusammen. In einer angeregten Diskussion wurde einstimmig entschieden, dass weder FirmenvertreterInnen noch Vertreter der Sozialpartner in diese

Expertengruppe einbezogen werden sollen. Die VertreterInnen sollten somit rein fachlich motiviert sein.

Die Gruppe nimmt eine beratende Funktion wahr (etwa im Sinne eines wissenschaftlichen Beirates) und tagt in unregelmäßigen Abständen in Wien. In zwei von uns moderierten Sitzungen hat diese Gruppe sowohl die Kriterienliste (inklusive K.O.-Kriterien) erstellt als auch die Gewichtung der Kriterien vorgenommen. Zur Vorbereitung dazu gab es ein einleitendes Referat über Funktion und Wirkungsweise von Lernplattformen.

Für die Zwecke der Untersuchung haben wir das zu evaluierende Produkt Lernplattform dabei folgendermaßen definiert:

Eine Lernplattform ist ein Werkzeug für die Organisation und Betreuung webbasierten Lernens.

Zur gemeinsamen Ausarbeitung der Kriterien haben wir die Delphi-Methode (Seeger 1979) gewählt, d.h. wir haben nach der Sammlung und Dokumentation die Ergebnisse der gesamten Gruppe rückgemeldet und sie um Streichungen, Ergänzungen etc. gebeten, diese Vorschläge wiederum in die Kriterienliste eingearbeitet, diese wiederum an die ExpertInnen rückgemeldet, etc.

Die Mindestkriterien dienten vor allem dazu, eine grobe Sichtung der am Markt befindlichen Systeme anzuleiten und reine Präsentations- Verwaltungs- und Authoring-Tools von Lernmanagement Systemen abzugrenzen. Wichtig dabei war es uns, dass sowohl didaktische Funktionen des interaktiven Online-Lernens, als auch administrative Funktionen integriert waren. Diese Mindestkriterien wurden als sogenannte K.O.-Kriterien angewendet, d.h. jene Systeme, die diese nicht erfüllten, wurden bei der nächsten Phase nicht mehr berücksichtigt.

Dazu wurden die Anbieter von Lernplattformen angeschrieben. Das bloße Nicht-Vorliegen von Informationen bedeutete vorerst nicht, dass das Produkt ausgeschieden würde. Lieferte allerdings der Hersteller nach einer angemessenen Nachfrist die Informationen nicht, schied das Produkt aus.

Die folgenden Kriterien wurden als **Mindestkriterien** (K.O.-Kriterien) angewendet:

1. österreichweite Rahmenlizenz:

Es muss eine österreichweite Rahmenlizenz angeboten werden.
Ausgeschlossen wurde damit ein Pay-Per-User-Lizenzmodell.

2. Webbasierende Lösung:

die auf einem Standardprotokoll (TCP/IP) und einem Standardwebbrowser basiert. Die AnwenderInnen können mittels beliebigem Webbrowser am Lernprozess teilnehmen. Das Lernmanagement sowie das Ergebnis einer Content-Generierung mittels Wizards bzw. Templates muss die Möglichkeit enthalten, mit einem Webbrowser ohne jegliche Erweiterung (Viewer und Plugins) bedienbar zu sein.

3. Administration:

Die Verwaltung von Usern sowie die Organisation (Erstellen, Erweitern, Löschen) von Content (Lerninhalte) und Kursen müssen möglich sein.

Das System muss die Authentisierung und Autorisierung (Rollen, Gruppen, Rechte) vom entsprechenden Zugriffs- und Rechtemanagement übernehmen können. Konkret muss das System LDAP (Light Weight Directory Access Protocol) fähig oder entsprechend anpassbar sein.

Das System muss rollenbasierend sein.

Das System muss multilingual sein und zumindest die deutsche und englische Sprache einschließen.

4. Kommunikation:

innerhalb der Rollen (Lernende, Lehrende) und

zwischen den Rollen (Lernende-Lehrende)

muss in elektronischer Form unterstützt werden.

5. käuflich erwerbbar / open source / public domain / LSP-Lösung

Das System sollte käuflich erwerbbar, open source, public domain oder eine Mietlösung sein.

Einige der Anbieter im eLearning Bereich bieten Lernplattformen an, allerdings nur auf Mietbasis auf deren eigenen Servern (LSP - Learning Service Provider). Eine LSP-Lösung wird für den post-sekundären und tertiären Bereich für möglich gehalten, für den Schulbereich jedoch ausgeschlossen.

Um für die zweite Sitzung mit den ExpertInnen eine Kriterienliste für die Qualitative Gewichtung und Summierung (QGS) vorbereiten zu können, haben wir bereits in der ersten Sitzung eine Liste mit „Dimensionen zur Beschreibung der Kriterien für die QGS“ vorgestellt. Diese Dimensionen, die noch keine Kriterien, sondern eher Überbegriffe der späteren Kriterien darstellten, wurden hinterfragt, diskutiert, ergänzt und gestrichen. So wurde uns klar, in welche Richtung die Wünsche der ExpertInnen und damit auch jene unserer Auftraggeber gingen.

Diese Vorgehensweise ermöglichte es uns, bereits zum zweiten Zusammentreffen eine Liste der Kriterien zur Qualitativen Gewichtung und Summierung mitzubringen, welche nur noch eine Feinabstimmung erforderte. So war es möglich, in nur zwei Sitzungen auch noch die Gewichtung der Kriterien für die QGS durchzuführen.

Die Kriterien für die QGS lassen sich unter den folgenden Kategorien zusammenfassen:

- Kommunikation, Kooperation & Kollaboration (5 Kriterien)
- Didaktik (7 Kriterien)
- Administration (3 Kriterien)
- Technik (12 Kriterien)

Unser Hauptproblem dabei war es, die Kriterien aus dem pädagogischen Bereich (Didaktik und Kommunikation) eine größere Gewichtung im Gesamtverfahren zukommen zu lassen. In der ersten Fassung gab es im Bereich der Technik und Administration viele Kriterien, die einen sehr feinen Detaillierungsgrad hatten. Dadurch bestand die Gefahr, dass diese Merkmale alleine durch ihre Menge das Evaluierungsverfahren dominierten. Nachdem wir einerseits im Technikbereich ähnliche Funktionen zusammenfassten und andererseits die pädagogischen Kriterien stärker unterteilten, erhielten wir ein relativ ausgeglichenes Gesamtbild: 12 didaktisch-kommunikative zu 15 technisch-administrative Kriterien.

Diese 27 Kriterien wurden der Expertengruppe zur individuellen Gewichtung vorgelegt (siehe das Formular dazu im Anhang). Es ergab sich die folgende Gewichtung:

Die gewichtete Kriterienliste

Legende:

- | | | |
|---|---|-----------------|
| * | = | äußerst wichtig |
| # | = | sehr wichtig |
| + | = | wichtig |
| | = | weniger wichtig |
| 0 | = | nicht wichtig |

Kommunikation, Kooperation und Kollaboration:

synchron	*
asynchron	*
Annotationen	#
Gruppenbildung durch Rollen	*
externe Kommunikationstools integrierbar	#

Didaktik:

Lässt verschiedene Lehr- und Lernmodelle zu	*
Interaktive Tests	#
Interaktive Übungen	#
Modularisierung von Lerninhalten	*
Feedback zum Lernfortschritt	#
Autorenfunktionen	+
Learning-flow-Management	#

Administration:

Tracking von Usern	#
Rechnungsverwaltung	+
Personalisierung	*

Technik:

Anpassbarkeit	*
Erweiterbarkeit	*
Skalierbarkeit	*
Distributierbarkeit	*
Entspricht Standards	*
Unterstützt Open-Source Betriebssysteme	#
Support	*
Dokumentation	#
Sicherheit bei Datenweiterreichung	*
Ressourcenbedarf	#
Österreichischer Vertriebspartner	#
Unterstützt Standardobjekttypen u. -formate	*

Die Aufgabe der ExpertInnen bestand darin, jedes Kriterium mit einem der oben stehenden Symbole zu bewerten. Die Symbole wurden verwendet um von vornherein eine Addition der Merkmale zu einer Gesamtzahl zu verhindern.

Es fällt auf, dass die 0 (nicht wichtig) nie vergeben wurde. Da es dieselben Personen waren, die die Kriterienliste erstellten und anschließend gewichteten, ist dies nur folgerichtig. Funktionen, die als unbedeutend eingeschätzt worden sind, sind gar nicht erst als Kriterien aufgestellt worden.

☉ Die Sichtung und Dokumentation der am Markt befindlichen Lernplattformen

Alle Anbieter von eLearning-Plattformen, auf die wir im Rahmen unserer Recherche gestoßen sind, wurden angeschrieben und von der Evaluation informiert. In diesem ersten Schreiben haben wir den Auftraggeber der Evaluation und die Evaluationsziele bekannt gegeben, kurz das Projektdesign skizziert und die Unternehmen aufgefordert, zu den Mindestkriterien Stellung zu nehmen.

Unsere Community-Website [W4] diente dabei als Informationsplattform, auf der wir die Mindestkriterien sowie weitere Informationen zum Projekt zur Verfügung stellten.

Die nachfolgenden Auszüge aus dem online geführten und öffentlich zugänglichen Projektstagebuch dokumentieren, wie schwierig es war, die gewünschten Informationen auch zu bekommen. Dass wir auf einen renommierten Auftraggeber (die österreichische Bundesregierung) und auf ein universitäres (Forschungs-)umfeld verweisen konnten, half nur teilweise. Nach einer Antwortfrist von 14 Tagen lagen aus der Liste der rund 130 angeschriebenen Unternehmen erst Stellungnahmen von 13 (!) Unternehmen vor. Vier davon konnten die Mindestkriterien nicht erfüllen.

Aufgrund der geringen Anzahl an Rückmeldungen verfassten wir zweimal Erinnerungsschreiben, die wir an alle auf den diversen Websites genannten Adressen sandten. In diesen Schreiben skizzierten wir nochmals kurz Ziel und Auftraggeber des Projektes, wiesen zur weiteren Information auf unsere Webseite hin und forderten zur Stellungnahme zu den Mindestkriterien auf. Einige Unternehmen (insbesondere die „big-player“ am Markt) haben wir, nachdem das 3-malige Anschreiben keine Reaktion gezeigt hatte, zusätzlich per Fax und / oder telefonisch kontaktiert.

Schließlich haben 49 Unternehmen zu den Mindestkriterien Stellung genommen. Davon konnten 41 diese Kriterien auch erfüllen und weiter im Evaluationsprozess verbleiben.

Wie dynamisch der e-Learning Markt ist, lässt sich auch am Verlauf dieser Recherche gut darstellen: Von den gut 130 Unternehmen, die wir recherchiert hatten, existierten nach den wenigen Wochen, in denen wir zur Stellungnahme zu den Mindestkriterien aufgefordert hatten, zwölf nicht mehr, bzw. war deren Internetseite nicht mehr existent. Vier andere Unternehmen wurden aufgekauft und deren Produkte umbenannt.

Die meisten Rückfragen zu den Mindestkriterien betrafen die Forderung nach einer Unterstützung der deutschen und englischen Sprache. Diese Forderung schien vor allem auf dem US-amerikanischen und kanadischen Markt noch nie aufgetaucht zu sein und so vergewisserten sich fast alle Unternehmen dieser Märkte, ob denn diese Forderung wirklich ein „K.O.-Kriterium“ darstelle.

Auch stellte die Forderung nach der LDAP-Fähigkeit ein Kriterium dar, das bei vielen Product-Consultants und Technikern der angesprochenen Unternehmen weiteren Erklärungsbedarf auslöste.

Im Anschluss an die Stellungnahme zu den Mindestkriterien haben wir die 40 [W5] noch in der Evaluation verbliebenen Unternehmen aufgefordert, zu den Kriterien für die Qualitative Gewichtung und Summierung Stellung zu nehmen. Die Liste der Kriterien sowie die Hinweise darüber, wie die Kriterien gewichtet werden, wurden wiederum auf der Communityseite zur Verfügung gestellt.

Dieses Mal hatten wir der Einladung zur Stellungnahme auch den Hinweis beigefügt, dass die Zuverlässigkeit, mit der unsere Anfragen bearbeitet würden, in die Evaluation mit einfließen werde. Zur vorgesehenen Deadline am 15. September lagen uns von den 40 eingeladenen Firmen, die Stellungnahmen von immerhin 24 Unternehmen vor.

Einige Stellungnahmen waren jedoch sehr spärlich formuliert und beschränkten sich auf ein einfaches „Ja“ bzw. „Nein“ zu den einzelnen Kriterien. Diese Unternehmen forderten wir nochmals auf, ihre Stellungnahmen zu präzisieren, um die Vergleichbarkeit unter den Produkten sicher zu stellen.

Auch für diese Phase hatten wir uns und den Unternehmen einen „endgültig letzten Termin“ gesetzt. Dieser wurde in persönlichen Mails und zusätzlichen Telefonaten den AnsprechpartnerInnen mitgeteilt und wiederum auf der Community-Seite veröffentlicht. Nach Ablauf dieser Nachfrist, fehlten uns immer noch die Stellungnahmen von vier Unternehmen. Immerhin drei davon hatten sich telefonisch gemeldet und uns zugesagt, so schnell wie möglich zu antworten.

Sehr viel Augenmerk haben wir auf eine präzise Ausformulierung der Stellungnahmen zu den QGS durch unsere Ansprechpartner verwendet. Dies stellte für uns einen erheblichen Aufwand dar, da alle Stellungnahmen durchgearbeitet und jedem Unternehmen explizite Fragen zu den gegebenen Antworten gestellt werden mussten. Dieser Aufwand ermöglichte es uns jedoch, die Bewertung der Kriterien für die QGS der einzelnen Produkte so fair wie möglich zu gestalten.

Auch hatten wir alle Verantwortlichen um die Zusendung eines Whitepapers zum Produkt gebeten, um möglichst viel aussagekräftige Zusatzinformationen zu erhalten. Dieser Bitte kamen alle Unternehmen nach, allerdings nur nach der ausdrücklichen Zusage unsererseits, diese ausführlichen Produktinformationen nicht an andere Firmen weiterzugeben („non disclosure agreement“). Insgesamt konnten so ca. 3100 Seiten an technischen Informationen zusammengetragen werden (94 Megabyte an PDF- und Word-Dokumenten).

Im Projektverlauf stellte sich heraus, dass wir eine zusätzliche Evaluationsschleife einbauen mussten, um so gründlich wie möglich zu bewerten. Diese betraf vor allem die Kriterien des

Punktes „Didaktik“, war aber auch notwendig, um das Gesamtbild der Lernplattformen abzurunden.

So forderten wir von allen Herstellern, deren Produkte die Mindestkriterien erfüllten, einen Testaccount an. Diese Bitte stellte einige UnternehmensrepräsentantInnen vor große Schwierigkeiten. Gerade die Qualitätssicherungssysteme der großen US-amerikanischen Firmen lassen es nicht zu, dass ein potentieller Käufer (oder Evaluator) das Produkt zu Gesicht bekommt, ohne dass ein Repräsentant der Firma zugegen ist. Dies hätte bedeutet, dass wir den freundlichen Einladungen zur Produktpräsentation quer durch Europa hätten Folge leisten müssen, oder, dass VerkäuferInnen die Produkte bei uns vor Ort präsentiert hätten. Beidem konnten wir nicht zustimmen, da dies eine Verzerrung zugunsten jener, die das Produkt persönlich präsentieren können, ergeben hätte. Nach langer Überzeugungsarbeit von Europa (zuerst von unserer Seite dann auch mit Hilfe der europäischen Niederlassungen) nach Amerika schafften es dann doch alle Unternehmen, einen solchen Account zur Verfügung zu stellen.

Diese zusätzliche Evaluationsschleife bedeutete einen nicht unbeträchtlichen Mehraufwand für uns. Dass sie sowohl wichtig als erfolgreich war, zeigen die Ergebnisse in Form einer qualitativen Beschreibung der Funktionen, die wir demnächst in Buchform präsentieren werden. Außerdem konnte durch die intensive Kommunikation mit dem Firmenrepräsentanten auch einige Missverständnisse, welche in den Stellungnahmen zu den QGS aufgetaucht waren, ausgeräumt werden.

☉ Die Erstellung der Shortlist

Aus den auf dem Markt befindlichen Angeboten wurden durch die oben dargestellte Vorgehensweise 16 Produkte¹ ausgewählt und zu einer Shortlist zusammengefasst. Innerhalb dieser Liste wurde jedoch keine Reihung vorgenommen. Unser Verzicht auf ein detailliertes Ranking hat mehrere Gründe:

- Das QGS-Verfahren liefert nicht immer eindeutige Ergebnisse. Zwar lässt sich sehr gut eine Gruppierung vornehmen, ein definitives Ranking zwischen einzelnen Produkten ist aber in einigen Fällen nur durch einen intensiven Vergleich der beiden Produkte möglich.
- Es wurde uns seitens einiger Firmen verboten ein definitives Ranking zu veröffentlichen. Zwar dürfen wir darüber informieren, ob das betreffende Produkt die Top 15 erreicht oder nicht erreicht hat; welchen Rang es dabei aber insgesamt einnimmt, diese differenzierte Aussage dürfen wir nicht publizieren.

Für die Zwecke unserer Untersuchung war diese Einschränkung aber weniger störend als es auf den ersten Blick erscheinen mag:

In der zweiten Phase werden die Top 15 weiter eingegrenzt (Top 5).

Die Gewichtung der Kriterien ist bei jeder Untersuchung – entsprechend den jeweiligen Wünschen und Zielen – vorzunehmen und beansprucht keine objektive Gültigkeit.

Mit einer guten qualitativen Beschreibung – die wir im kommenden Buch veröffentlichen werden - ermöglichen wir es Interessenten, diese Gewichtungen selbst vorzunehmen und damit ihr eigenes Ranking zu erstellen.

Nachfolgend finden Sie das Formular, welches wir zur Bewertung der Lernplattformen entwickelt und eingesetzt haben. Die Bewertungen für die Plattformen A, B und C haben wir beispielhaft eingesetzt.

¹ 16 Produkte deshalb, da sich im Laufe der Evaluation herausstellte, dass die Lernplattform "Campus 2000" der ibis acam G.m.b.H auf dem "IBT-Server" der time4you G.m.b.H beruht. So wurde das letztgenannte Produkt noch in die Evaluation aufgenommen. Die nachfolgende Zusammenstellung berücksichtigt nur mehr den IBT-Server und ist also wieder eine Liste der Top 15.

Kriterium	Gewichtung	Plattform A				Plattform B				Plattform C			
Kommunikation													
synchron	#	+				+							
asynchron	*	*				#				*			
Annotationen	#	0				0				+			
Gruppenbildung durch Rollen	*	*				*				*			
Externe Komm.tools	#					+				#			
Didaktik													
Versch. Lehr- u. Lernmodelle	*	*				#				*			
Interaktive Tests	#	#				+				#			
Interaktive Übungen	*	*				*				*			
Modularisierung	*	*				#				*			
Feedback zum Lernfortschritt	#	#				0				#			
Autorenfunktionen	#	#				#				#			
Learning-flow-Management	#	#				#				#			
Administration													
Tracking von Usern	+	+								+			
Rechnungsverwaltung	+	+				+							
Personalisierung	#	+								#			
Technik													
Anpassbarkeit	*	*								*			
Erweiterbarkeit	*					0				*			
Skalierbarkeit	#	#				#				#			
Entspricht Standards	*					#				*			
open-source	#	#				#							
Support	*	*				*				*			
Dokumentation	#	#				#				#			
Datensicherheit	#	#								#			
Ressourcenbedarf	#	#				#				#			
Österr. Vertriebspartner	#	#				0				0			
Standardobjekttypen	*	*				*				*			
Anzahl der Gewichtungen		*	#	+		*	#	+		*	#	+	
		8	10	4	3	4	10	4	4	10	10	2	3

Wir haben die Produkte jener Unternehmen, die zu den Kriterien der Qualitativen Gewichtung und Summierung Stellung genommen haben, nach der Bewertung insgesamt in vier Cluster eingeteilt.

Die Ergebnisse

Die nachfolgende Reihung innerhalb der genannten Listen ist alphabetisch und stellt keine Wertung dar.

A. Kriterien nicht ausreichend erfüllt

Diese Lernplattformen erfüllen wohl die Mindestkriterien, jedoch die Kriterien der QGS nicht in ausreichendem Maße.

B. Kriterien erfüllt

Diese Lernplattformen erfüllen sowohl die Mindestkriterien als auch die Kriterien der QGS. Die Kriterien der QGS werden jedoch nicht in der Qualität erfüllt, wie sie notwendig gewesen wäre, um in die Shortlist aufgenommen zu werden.

C. Kriterien gut erfüllt, das Produkt ist jedoch nicht deutschsprachig²

	Produkt:	Firma:
1	eCollege	eCollege
2	Nebo	Learnframe

Diese Lernplattformen würden in der Shortlist vertreten sein, denn sowohl die Mindestkriterien als auch jene der QGS werden in ausreichender Qualität erfüllt. Jedoch sind die Produkte nicht deutschsprachig und können daher in der von uns durchgeführten Studie nicht berücksichtigt werden.

² Nach den Angaben der Hersteller würde das Produkt bei Vertragsabschluss binnen 3 Monaten lokalisiert werden.

D. Kriterien gut erfüllt (= Shortlist)

Diese Lernplattformen stellen die Shortlist dar. Es sind dies jene 15 Produkte, welche die Kriterien der QGS am besten der Gewichtung des ExpertInnenteams entsprechend erfüllen. Diese Produkte werden wir demnächst in einer eigenen Publikation ausführlich beschreiben und in ihrer Funktionalität mit Bildschirmausdrucken („screenshots“) dokumentieren.

	Produkt:	Firma:
1	Blackboard	Blackboard
2	Clix	imc G.m.b.H
3	DLS	ets G.m.b.H
4	Docent	Docent Inc.
5	eLS	Hyperwave AG
6	IBT Server	Time4You GmbH
7	iLearning	Oracle
8	Ilias	Uni Köln
9	Learning Space	Lotus (IBM)
10	Learntone	Sun Microsystems
11	Saba Learning Enterprise	Saba
12	Sitos	Bitmedia
13	Thinktanx	Viviance
14	Top Class	WBT Systems
15	WebCT	WebCT

Wie bereits ausgeführt stellt die Reihung der Shortlist keine Wertung dar, sondern ist bloß eine nach Produktnamen alphabetisch geordnete Liste.

Darüber hinaus führen wir noch die Add-on Produkte (Virtual-Classroom Produkte und Assessment Tools) an, welche sich an der Evaluation beteiligt hatten, da sie die Lernplattformen in Teilbereichen sinnvoll ergänzen. Wir hatten die jeweiligen Hersteller- bzw. Vertriebsfirmen davon in Kenntnis gesetzt, dass die Evaluanden dieses Projektes ausschließlich asynchrone Lernplattformen sind, weshalb ihre Produkte in der Shortlist keine Erwähnung finden können.

a. Synchroner Lernplattformen:

Eine zusätzliche Gruppe bilden die synchronen Lernplattformen (Virtual-Classroom-Tools), die ebenfalls die Mindestkriterien erfüllt haben und als Add-On-Produkte zu den oben angeführten Learning-Management-Systemen zu sehen sind.

Virtual-Classroom-Tools unterstützen das synchrone Lernen mit Funktionen wie Whiteboard, Application-Sharing, Ad-Hoc-Fragen, gemeinsames Surfen, File-Sharing, etc. für 1:1 und 1:n Verbindungen. Über Schnittstellen können die Produkte auch in die in der Shortlist vertretenen asynchronen Lernplattformen integriert werden.

	Produkt:	Firma:
1	Centra Symposium	Centra Inc.
2	Interwise ECP	Interwise
3	Learnlinc	Mentergy

b. Assessment Tools:

Assessment Tools erlauben das Erstellen und Verwalten von Quizzes, Tests und Befragungen über das Internet. Über Schnittstellen können die Produkte auch in Lernplattformen integriert werden, die selbst nicht über ein entsprechendes Tool verfügen.

	Produkt:	Firma:
1	Perception	Question Mark
2	Cyberline	Tertia Edusoft

Zusammenfassung

Obwohl das Verfahren der Qualitativen Gewichtung und Summierung relativ komplex und aufwendig ist, hat es sich gezeigt, dass es auch für reale Situationen außerhalb eines universitären Umfelds durchaus praktikabel ist. Vorbedingung ist allerdings die Bildung eines ExpertInnen-Gremiums, das die Kriterien erstellt, gewichtet und den gesamten Prozess der Evaluierung begleitet.

Weiters hat sich gezeigt, dass neben der eigentlichen Evaluierungstätigkeit (Prüfen und Bewerten der Evaluanden an Hand der Kriterienlisten) ein hoher Kommunikationsbedarf notwendig ist. Damit werden nicht nur mögliche Missverständnisse bereinigt, sondern auch die Vertreter der zu evaluierenden Produkte in den Evaluierungsprozess mit einbezogen. Besonders in Beispielen wie dem unseren, in dem es keine einheitliche internationale Agentur gibt, die eine für alle Firmenvertreter verbindliche gemeinsame Evaluierung durchführen würde, ist ein ständiger Kommunikationsfluss notwendig.

Allerdings hat sich auch gezeigt, dass Firmen eine individuelle Kommunikation (mündliches Gespräch) gegenüber öffentlichen Aussagen (z.B. über Webseiten, auf denen wir Firmen zu Stellungnahmen eingeladen haben) vorziehen.

Es bleibt zu hoffen, dass unsere Methode breite Akzeptanz findet und mithilft ein Umdenken bezüglich der Beteiligung an dem Evaluierungsverfahren auch bei den „big players“ einzuleiten. Insbesondere dann, wenn klar wird, dass sowohl die aufgestellten Kriterien als auch ihre Gewichtung durchaus subjektiven Charakter haben, für jede Untersuchung und Interessengruppierung neu erstellt werden müssen und es daher keine „objektiv und ewig wahren Ergebnisse“ gibt.

Außerdem ist der Markt einer ständigen Bewegung unterworfen (neue Produkte, neue Versionen bestehender Produkte), sodass solch eine Evaluierung nur eine Momentaufnahme darstellen kann. Allerdings planen wir den während der Studie hergestellten guten Kontakt mit den FirmenvertreterInnen zu nutzen, um ständige Aktualisierungen unserer Ergebnisse vorzunehmen.

Web Referenzen:

[W1]: EASA (European Academic Software Award): <http://www.easa.org>

[W2]: MeDiDa Prix (Mediendidaktischer Hochschulpreis): <http://www.medidaprix.org>

[W3]: Projekttagbuch: <http://www.virtual-learning.at/projekttagbuch.htm>

[W4]: Virtual-Learning Community Website: **<http://www.virtual-learning.at/community>** (Dieser Website stellt die Methodik und den Prozess des Evaluierungsverfahrens dar und berichtet kontinuierlich über alle Ergebnisse der Untersuchung).

[W5]: Linkliste der 40 Lernplattformen, die die Mindestkriterien erfüllt haben:
http://iol3.uibk.ac.at/virtualllearning/Ergebnisse/Links_Mindestkriterien

Literatur Referenzen:

Baumgartner, P. (2001). Webbasierte Lernumgebungen - neue Ansätze zum Lernen. In: Politikunterricht im Informationszeitalter. Hg. G. Weißeno. Schwalbach/Ts., Wochenschau Verlag: 90-104.

Baumgartner, P. 1998. Hochschuldidaktik als Organisationsentwicklung. Neue Medien als Katalysator einer anderen Sichtweise. In: media@uni - multi.media? Entwicklung - Gestaltung - Evaluation neuer Medien, Hg. von M. Hauff. Münster: Waxmann. 57- 62.

Baumgartner, P. 1997. Didaktische Anforderungen an (multimediale) Lernsoftware. In: Information und Lernen mit Multimedia, Hg. von L. J. Issing und P. Klimsa. 2. überarbeitete . Aufl. Weinheim: Psychologie-Verl.-Union. 241-252.

Baumgartner, P. and S. Frank (2000). Der Mediendidaktische Hochschulpreis (MeDiDa-Prix) - Idee und Realisierung. In: Campus 2000 - Lernen in neuen Organisationsformen. F. Scheuerman. Münster, Waxmann: 63-81.

Baumgartner, P. and S. Payr (1997). Methods and practice of software evaluation: The case of the European Academic Software Award (EASA). In: Proceedings of ED-MEDIA 97 - World Conference on Educational Multimedia and Hypermedia. Charlottesville, AACE: 44-50.

- Biermann, H. 1994. Lehren und Lernen mit Computern. In: Lehren und Lernen im Umfeld neuer Technologie: Reflexionen vor Ort, Hg. von J. Petersen und G.-B. Reiner. Frankfurt/Main: Peter Lang. 123-141.
- Doll, C. A. 1987. Evaluating Educational Software. Chicago/London: American Library Association.
- Fricke, R. 1995. Evaluation von Multimedia. In: Information und Lernen mit Multimedia, Hg. von L. J. Issing und P. Klimsa. Weinheim: Psychologie-Verl.-Union. 401-413.
- Scriven, M. 1991. Evaluation Thesaurus. 4. Aufl. Newbury Park: SAGE.
- Seeger, T. 1979. Die Delphi-Methode: Expertenbefragung zwischen Prognose u. Gruppenmeinungsbildungsprozessen. Freiburg: Hochschul-Verlag.
- Thomé, D. 1988. Kriterien zur Bewertung von Lernsoftware. Heidelberg: Hüthig.