

Σύγκριση Εμπειρικών Μεθοδολογιών Αξιολόγησης και Μεθοδολογιών Βασισμένων σε Ειδικούς στην Περίπτωση ενός CBL Περιβάλλοντος: Η Εμπειρία του «Ορέστη».

Καρούλης, Αθ., Λιώτσιος, Κων., Δημητριάδης, Στ., Πομπόρτσης, Αν., και
Τσουκαλάς, Ι.

Τμήμα Πληροφορικής – Αριστοτέλειο Παν/μιο Θεσσαλονίκης
PO Box 888 – 54124 Θεσσαλονίκη
[\[karoulis, liotsios, sdemetri, apombo, tsoukala}@csd.auth.gr](mailto:{karoulis, liotsios, sdemetri, apombo, tsoukala}@csd.auth.gr)

ΠΕΡΙΛΗΨΗ

Στην εργασία αυτή συγκρίνονται περισσότερες μεθοδολογίες αξιολόγησης διασυνδέσεων, οι οποίες εφαρμόστηκαν στην περίπτωση ενός Περιβάλλοντος Μάθησης Βασισμένου σε Υπολογιστή (Computer Based Learning – CBL) κατά τη διάρκεια μιας μακροχρόνιας έρευνας σε τρία Ευρωπαϊκά κράτη, Ελλάδα, Γερμανία και Ολλανδία και στα πλαίσια ενός χρηματοδοτούμενου προγράμματος Leonardo da Vinci. Στη μελετώμενη περίπτωση ένας συνδυασμός μεθοδολογιών αξιολόγησης βασισμένων σε ειδικούς και βασισμένων σε χρήστες (εμπειρικές) εφαρμόστηκε στο περιβάλλον και ο «ιδανικός» συνδυασμός διερευνήθηκε, με σκοπό την ελαχιστοποίηση των πόρων και τη μεγιστοποίηση των αποτελεσμάτων κατά την αξιολόγηση του λογισμικού. Τα αποτελέσματα της μελέτης έδειξαν ότι και οι δύο προσεγγίσεις – βασισμένες σε ειδικούς και εμπειρικές – απέδωσαν ικανοποιητικά, αν και οι εμπειρικές μέθοδοι έδειξαν να είναι προτιμητέες. Έτσι, η έρευνα αυτή καταλήγει στην πρόταση μιας συνδυαστικής προσέγγισης αξιολόγησης βασισμένης σε ειδικούς κατά τα πρώτα στάδια σχεδιασμού και κατασκευής του προϊόντος με μια παρατηρητική αξιολόγηση σε πιο προχωρημένα στάδια που φαίνεται ότι παρείχε τον καλύτερο συντελεστή απόδοσης / κόστους.

ABSTRACT

In this paper several interface evaluation methods are compared, applied in the case of a Computer Based Learning (CBL) environment, during a longitudinal study performed in three European countries, Greece, Germany, and Holland, and within the framework of a EC funded Leonardo da Vinci program. In the case studied, a combination of expert-based approaches with usability testing is dealt with, and the «ideal» combination is investigated, in order to consume the fewest possible resources by achieving maximum outcome for the assessment of the software. The results of this study show that both approaches, expert-based and empirical, have been found to perform adequately, although the empirical methodologies are always preferable. So, this study concludes by proposing a combination of an expert-based approach during the early stages of the design cycle with an observational evaluation, performed by the instructors and/or the designers, that proved to provide the best cost/performance factor.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: CBL περιβάλλοντα, αξιολόγηση με ειδικούς, εμπειρική αξιολόγηση

ΕΙΣΑΓΩΓΗ

Η έρευνα αυτή αφορά την αξιολόγηση ευχρηστίας (usability evaluation) ενός CBL λογισμικού, ονόματι “Ορέστης”, το οποίο παρήχθη στο εργαστήριο πολυμέσων του τμήματος πληροφορικής

του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης. Τρία ευρωπαϊκά κράτη συμμετείχαν στην αξιολόγησή του, εφαρμόζοντας δέκα συνεδρίες αξιολόγησης από έξι διαφορετικές μεθόδους αξιολόγησης που ανήκουν σε δύο μεθοδολογίες, τις εμπειρικές αξιολογήσεις και τις αξιολογήσεις βασισμένες σε ειδικούς. Ο σκοπός της μελέτης είναι να διερευνήσει την εφαρμογή περισσότερων μεθόδων αξιολόγησης πάνω στο ίδιο λογισμικό. Αρχικά παρέχονται γενικές πληροφορίες σχετικά με τις μεθοδολογίες και τις μεθόδους αξιολόγησης που χρησιμοποιήθηκαν, και ακολούθως τα αποτελέσματα από την εφαρμογή τους στο λογισμικό που παρουσιάζεται εδώ. Οι βασικές μεθοδολογίες αξιολόγησης που ερευνώνται σ' αυτή την εργασία είναι οι αξιολογήσεις ειδικών και η εμπειρική (βασισμένη στους χρήστες). Στις προσεγγίσεις που βασίζονται σε ειδικούς, ζητείται από μια διεπιστημονική ομάδα ειδικών (ειδικοί του τομέα επικοινωνίας ανθρώπου – υπολογιστή (Human-Computer Interaction - HCI), ειδικοί γνωστικών επιστημών και ειδικοί στον υπό μελέτη τομέα) να κρίνουν την διασύνδεση και να επισημάνουν τα προβλήματα που οι εν δυνάμει χρήστες πιθανώς να συναντήσουν κατά την διάρκεια της ενασχόλησής τους με αυτή. Σε μία εμπειρική προσέγγιση, οι χρήστες παρατηρούνται κατά την διάρκεια της ενασχόλησής τους με την διασύνδεση, ή τους δίνονται ερωτηματολόγια και/ή τους παίρνεται συνέντευξη. Οι βασισμένες σε ειδικούς μέθοδοι υποστηρίζεται ότι είναι εφαρμόσιμες από τα αρχικά στάδια σχεδιασμού της διασύνδεσης και καταναλώνουν λιγότερους πόρους, όμως δεν μπορούν να προβλέψουν επακριβώς τις αντιδράσεις των χρηστών. Από την άλλη, οι εμπειρικές μεθοδολογίες απαιτούν σημαντικούς πόρους και μπορούν να εφαρμοστούν μόνο μετά από ένα συγκεκριμένο σημείο ωριμότητας του λογισμικού, όμως μπορούν να δώσουν πολύτιμα ποιοτικά αποτελέσματα για να βελτιώσουν το αντικείμενο της αξιολόγησης.

Έγιναν τέσσερις συνεδρίες αξιολογήσεων με ειδικούς: η τροποποιημένη Γνωστική Γραφική Διάσχιση (Cognitive Graphical Jogthrough) (Karoulis et al., 2000), μία φαινομενογραφική (phenomenographical) προσέγγιση (Marton, 1988), μία επίσημη ανασκόπηση ειδικών (formal expert review) (Nielsen, 1993) και μία καταμετρική αξιολόγηση ειδικών βασισμένη στο ερωτηματολόγιο QUIS (Shneiderman, 1998). Επίσης έγιναν έξι εμπειρικές συνεδρίες: δύο παρατηρητικές, τρεις καταμετρητικές (βασισμένες σε ερωτηματολόγια, η μία είναι με το QUIS) και μία σε εργαστήριο ευχρηστίας. Αυτές οι μέθοδοι παρουσιάζονται σε συντομία, όμως το βασικό σημείο σ' αυτή την εργασία είναι η σύγκριση της αποδοτικότητας κάθε χρησιμοποιηθείσας μεθόδου μέσα στο πλαίσιο του τομέα της αλληλεπίδρασης ανθρώπου-υπολογιστή (HCI), καθώς επίσης και τα παρατηρηθέντα πλεονεκτήματα και μειονεκτήματα κάθε μεθόδου που εφαρμόστηκε στο ίδιο κομμάτι εκπαιδευτικού λογισμικού.

ΥΠΟΒΑΘΡΟ

Ένα Περιβάλλον Μάθησης βασισμένο σε Υπολογιστή (CBL) είναι ένα λογισμικό που καλύπτει γνωστικά ένα συγκεκριμένο πεδίο και παρέχει στον εκπαιδευόμενο όλα τα μέσα για να αποκτήσει γνώση πάνω στο πεδίο αυτό. Αυτός ο ορισμός υπαινίσσεται δύο προϋποθέσεις: η γνωστική κάλυψη και η παρουσίαση του υπό μελέτη τομέα είναι πλήρης (θεωρία, ασκήσεις, προσομοιώσεις) και ο εκπαιδευόμενος μπορεί να αλληλεπιδράσει με το αντικείμενο μελέτης (αλληλεπίδραση με το λογισμικό, με τα πολυμεσικά στοιχεία κλπ), με άλλα λόγια, υπάρχει ένα κατάλληλο κανάλι επικοινωνίας για να μπορέσει ο εκπαιδευόμενος να αποκτήσει την προσφερόμενη γνώση. Τα CBL περιβάλλοντα συγχέονται συχνά με τα πολυμεσικά περιβάλλοντα, με τις διάφορες διαδικτυακές εκπαιδευτικές πηγές, ή με τα εξ αποστάσεως εκπαιδευτικά περιβάλλοντα. Όμως υπάρχει μια μεγάλη διαφορά η οποία διακρίνει αυτά τα είδη περιβαλλόντων από τα εξ αποστάσεως εκπαιδευτικά. Στα CBL περιβάλλοντα υπάρχει περιορισμένη μόνο υποστήριξη για τον εκπαιδευόμενο, αρκετή όμως για να διευκολύνει την μεταβίβαση γνώσης (πχ σαν λειτουργία παροχής βοήθειας). Αυτά τα περιβάλλοντα προϋποθέτουν την παρουσία

εκπαιδευτή και ο σκοπός τους είναι να τον βοηθήσουν και να τον υποστηρίξουν στην δουλειά του. Εναλλακτικά μπορούν να χρησιμοποιηθούν σε γνωστικό περιβάλλον με το οποίο ο χρήστης είναι ήδη εξοικειωμένος, με σκοπό να εμπλουτίσουν την γνώση του – πχ μία βάση γνώσης (knowledge database) για τους εργαζόμενους μιας εταιρείας. Το μεγάλο πλεονέκτημα τέτοιων περιβαλλόντων είναι η σχετικά εύκολη και χαμηλού κόστους παραγωγή τους, καθώς το μόνο που έχουν να κάνουν είναι να διατηρούν μιας βασικής μορφής υποστηρικτική δομή προς τον εκπαιδευόμενο. Τα CBL περιβάλλοντα περιέχουν εξ ορισμού την πολυμεσικού προσανατολισμού σχεδίαση και μπορούν να βρίσκονται στο διαδίκτυο ή σε μορφή CD. Επί πλέον, το εκπαιδευτικό λογισμικό γενικά πρέπει να υπακούει στις βασικές εκπαιδευτικές προσεγγίσεις και θεωρίες, οι οποίες πρέπει να λαμβάνονται σοβαρά υπ' όψη κατά την φάση της σχεδίασης και της υλοποίησής του.

Υπάρχει ανεπαρκής μαρτυρία για την αξιολόγηση CBL περιβαλλόντων. Ειδικότερα, δεν υπάρχει σχεδόν καμία αναφορά για σύγκριση μεταξύ αξιολογήσεων βασισμένων σε ειδικούς και εμπειρικών για τέτοια περιβάλλοντα. Ο σκοπός αυτής της εργασίας δεν είναι συνεπώς να παρουσιαστούν τα αποτελέσματα αξιολόγησης του λογισμικού, αλλά περισσότερο να παρουσιαστούν τα συγκριτικά αποτελέσματα των μεθόδων αξιολόγησης που εφαρμόστηκαν. Η κύρια ερώτηση είναι κατά πόσον οι μεθοδολογίες που βασίζονται σε ειδικούς αποδίδουν ικανοποιητικά, εφορμώντας από την παραδοχή ότι οι εμπειρικές προσεγγίσεις δίνουν πάντα τα καλύτερα αποτελέσματα. Το δεύτερο ερώτημα είναι, ποια από τις μεθόδους που εφαρμόστηκαν απέδωσε συγκριτικά καλύτερα, μιας και είναι ασυνήθιστο να εφαρμόζονται πολλές διαφορετικές μέθοδοι αξιολόγησης για να αποτιμήσουν το ίδιο λογισμικό. Παρ' όλα αυτά, μιας και αυτό το πρόγραμμα ήταν μία διεθνής συνεργασία, κάθε συμμετέχων φορέας ήταν ευέλικτος στο να εφαρμόσει τις δικές του μεθόδους αξιολόγησης. Έτσι τα αποτελέσματα συνοψίζουν και συγκρίνουν ένα πλήρες φάσμα μεθόδων προσδιορίζοντας την σχετική δυναμική του καθενός.

ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΛΟΓΙΣΜΙΚΟΥ

Το CBL περιβάλλον «Ορέστης» είναι μέρος του προγράμματος Leonardo da Vinci «Βελτίωση ικανοτήτων και δεξιοτήτων ανηλίκων παραβατών και ανηλίκων σε κίνδυνο» και ήταν μία συνεργασία μεταξύ αρκετών οργανισμών σε τέσσερις ευρωπαϊκές χώρες Ελλάδα, Γερμανία, Ολλανδία και Ουγγαρία. Δυστυχώς, στην πορεία του έργου ουδεμία ανάδραση υπήρξε από ουγγρικής πλευράς (σχετική με την παρούσα εργασία), συνεπώς εν συνεχεία ο εταίρος αυτός «αποσιωπάται». Το κυρίως CBL λογισμικό σχεδιάστηκε και κατασκευάστηκε στο εργαστήριο πολυμέσων του ΑΠΘ στην Ελλάδα. Τέσσερις πανομοιότυπες εκδοχές παράχθηκαν στις τέσσερις γλώσσες Ελληνικά, Γερμανικά, Ολλανδικά και Ουγγρικά. Το CD-Rom “Ορέστης” είναι ένα πολυμεσικό CBL περιβάλλον που σκοπεύει τη διδασκαλία βασικών και προχωρημένων δεξιοτήτων για τη χρήση και τη συντήρηση ενός φωτοτυπικού καταστήματος σε νεαρούς παραβάτες του νόμου. Ολόκληρο το πρόγραμμα αποτελείται από τρεις φάσεις: κατασκευή του CD-Rom, εφαρμογή της εκπαιδευτικής μεθοδολογίας στις συμμετέχουσες χώρες και αξιολόγηση του αποτελέσματος.

Το CD-Rom “Ορέστης” προσομοιώνει ένα κατάστημα φωτοτυπιών με τα βασικά μηχανήματα (ασπρόμαυρο φωτοαντιγραφικό, φωτοτυπικό, μηχανή δεσίματος σπιράλ, θερμοκολλητική μηχανή και πλαστικοποιητική μηχανή), καθώς και τα αναλώσιμα υλικά (χαρτιά, διαφάνειες, τόνερ, γραφίτης κλπ). Η γνώση που προσφέρθηκε δομήθηκε σε θεωρία, ασκήσεις και πρακτική εξάσκηση. Η θεωρία προσφέρεται με κείμενο εμπλουτισμένο με πολυμεσικά στοιχεία, όπως εικόνες, ήχο και βίντεο. Μέσω μιας σειράς «κεφαλαίων» και «παραγράφων» παρέχεται στο σπουδαστή η σχετική θεωρία για κάθε θέμα. Υπάρχουν επίσης τριών ειδών ασκήσεις

ΑΞΙΟΛΟΓΗΣΗ ΤΗΣ ΔΙΑΣΥΝΔΕΣΗΣ ΧΡΗΣΤΗ

Η αξιολόγηση της διασύνδεσης ενός συστήματος λογισμικού, είναι μία διαδικασία που αποσκοπεί στο να αναγνωρίσει και να προτείνει λύσεις στα προβλήματα ευχρηστίας που προκαλούνται από τον συγκεκριμένο σχεδιασμό του λογισμικού. Ο όρος «αξιολόγηση» αναφέρεται γενικά στην διαδικασία «συλλογής δεδομένων σχετικά με την ευχρηστία ενός σχεδιασμού ή προϊόντος για μία ορισμένη ομάδα χρηστών, για μια συγκεκριμένη δραστηριότητα μέσα σε ένα καθορισμένο περιβάλλον ή πλαίσιο εργασίας» (Preece et al., 1994, σ.602). Ο κύριος σκοπός μιας αξιολόγησης διασύνδεσης, είναι, όπως ήδη αναφέρθηκε, να ανακαλύψει προβλήματα ευχρηστίας. Σαν πρόβλημα ευχρηστίας μπορεί να οριστεί «οτιδήποτε παρεμβαίνει στην ικανότητα του χρήστη να ολοκληρώσει αποδοτικά και αποτελεσματικά μια διεργασία» (Karat et al., 1992). Η αξιολόγηση του σχεδιασμού της διασύνδεσης χρήστη είναι σημαντικής σημασίας στο συνολικό πλάνο της αξιολόγησης του λογισμικού για δύο βασικούς λόγους. Πρώτον επειδή αφορά επακριβώς το τμήμα του λογισμικού που επιτρέπει στους χρήστες να μεταβιβάσουν τις οδηγίες τους στο μηχανήμα.

Σύμφωνα με την ταξινόμηση του Scriven (1976), υπάρχουν δύο κύριες κατηγορίες αξιολόγησης: η διαμορφωτική (ή προσαρμοστική – formative) και η αθροιστική (ή αποτιμητική – summative). Η πρώτη διεξάγεται κατά την διάρκεια της σχεδιαστικής και κατασκευαστικής φάσης, ενώ η δεύτερη διεξάγεται αφού το προϊόν έχει φτάσει στον τελικό χρήστη. Τα αποτελέσματα και συμπεράσματα της πρώτης χρησιμοποιούνται κυρίως για την αποσφαλμάτωση και για την βελτίωση των χαρακτηριστικών της διασύνδεσης (ανιχνεύοντας προβλήματα και αδυναμίες), ενώ τα αποτελέσματα της τελευταίας χρησιμοποιούνται για να βελτιώσουν την διασύνδεση σαν σύνολο και να εξυπηρετήσουν περισσότερες ανάγκες του χρήστη σε μία μελλοντική αναβάθμιση.

Οι Benyon et al. (1990) διακρίνουν τέσσερις κύριες μεθοδολογίες αξιολόγησης:

- *Αναλυτική αξιολόγηση.* Αυτή χρησιμοποιεί μία τυπική ή μη τυπική περιγραφή της διασύνδεσης ώστε να προβλέψει την συμπεριφορά των χρηστών από την άποψη φυσιοκινητικών και γνωστικών λειτουργιών που πρέπει να εκτελεστούν.
- *Αξιολόγηση ειδικών.* Ζητείται από ειδικούς να κρίνουν το σύστημα και να αναγνωρίσουν τα πιθανά προβλήματα ευχρηστίας, παίρνοντας τον ρόλο λιγότερο έμπειρων χρηστών.
- *Εμπειρική αξιολόγηση.* Σκοπός της είναι να συγκεντρώνει δεδομένα για την συμπεριφορά του χρήστη καθώς χρησιμοποιεί το σύστημα (παρατηρητική αξιολόγηση), ή επιστρατεύοντας τη χρήση συνεντεύξεων ή ερωτηματολογίων με σκοπό την εκμείευση της υποκειμενικής άποψης των χρηστών και την κατανόηση από αυτούς της διασύνδεσης (καταμετρητική αξιολόγηση).
- *Πειραματική αξιολόγηση.* Ο αξιολογητής μπορεί να χειριστεί ένα αριθμό παραγόντων που συνδέονται με τον σχεδιασμό της διασύνδεσης και να μελετήσει τις επιδράσεις τους σε ποικίλες όψεις της επίδοσης του χρήστη.

ΕΜΠΕΙΡΙΚΕΣ ΑΞΙΟΛΟΓΗΣΕΙΣ ΚΑΙ ΑΞΙΟΛΟΓΗΣΕΙΣ ΒΑΣΙΣΜΕΝΕΣ ΣΕ ΕΙΔΙΚΟΥΣ.

Οι περισσότερες εφαρμοζόμενες μεθοδολογίες είναι οι αξιολογήσεις ειδικών και οι εμπειρικές (βασισμένες σε χρήστες). Οι αξιολογήσεις ειδικών είναι μία σχετικά φθηνή και αποδοτική διαμορφωτική μεθοδολογία αξιολόγησης που εφαρμόζεται ακόμη και σε πρωτότυπα συστήματα, ή σχεδιαστικές προδιαγραφές μέχρι το σχεδόν «έτοιμο για παράδοση» προϊόν. Η κυρίως ιδέα είναι να παρουσιάσει τις διεργασίες που υποστηρίζονται από την διασύνδεση σε μία διεπιστημονική ομάδα ειδικών που θα πάρουν την θέση μελλοντικών χρηστών και θα προσπαθήσουν να αναγνωρίσουν πιθανές ανεπάρκειες στον σχεδιασμό της διασύνδεσης. Ωστόσο, σύμφωνα με τους Lewis & Rieman (1994) «δεν μπορείς να γνωρίζεις πόσο καλή ή κακή πρόκειται να είναι η

διασύνδεσή σου αν δεν βάλεις ανθρώπους να την χρησιμοποιήσουν». Αυτή η φράση εκφράζει την ευρεία πίστη πως ο εμπειρικός έλεγχος είναι αναπόφευκτος προκειμένου να αξιολογηθεί σωστά μία διασύνδεση. Γιατί λοιπόν δεν χρησιμοποιούμε μόνο εμπειρικές αξιολογήσεις, αλλά ερευνούμε επίσης και άλλες προσεγγίσεις; Ένα σημαντικό πλεονέκτημα αυτών των μεθόδων είναι ότι μπορούν να εφαρμοσθούν πολύ νωρίς στον κύκλο σχεδίασης, ακόμη και σε μακέτες επί χάρτου. Η εμπειρία του ειδικού του επιτρέπει να καταλάβει την λειτουργικότητα του υπό κατασκευή συστήματος, ακόμη και αν στερείται της γενικής εικόνας του προϊόντος. Μια πρώτη ματιά στα βασικά χαρακτηριστικά είναι συνήθως αρκετή για έναν ειδικό. Από την άλλη, οι αξιολογήσεις χρηστών μπορούν να εφαρμοσθούν μόνο αφού το προϊόν φθάσει σε ένα συγκεκριμένο επίπεδο ολοκλήρωσης. Οι αξιολογήσεις ειδικών ξεπερνούν αυτά τα εμπόδια αντικαθιστώντας τους χρήστες με ειδικούς, αν και δίνονται ποιοτικά φτωχότερα αποτελέσματα, αλλά με πολύ μικρότερο κόστος και δυσκολία στην διεξαγωγή, δηλαδή με ένα πολύ καλύτερο συντελεστή απόδοσης/κόστους.

ΕΦΑΡΜΟΓΗ ΣΤΗΝ ΠΡΑΞΗ

Μία μακροχρόνια μελέτη έγινε σε τρεις Ευρωπαϊκές χώρες, Ελλάδα, Γερμανία και Ολλανδία. Δυστυχώς δεν υπήρξε καμία αναφορά από την Ουγγαρία. Η παρουσίαση εδώ βασίζεται σε μία τελική έκθεση από την Γερμανία που αφορά την εφαρμογή σε μια φυλακή (Report 1, 2000), μία τελική έκθεση από την Ολλανδία (Report 2, 2000) που αφορά την εφαρμογή σε δύο φυλακές και δύο εφαρμογές που έγιναν στην Ελλάδα από τους συγγραφείς.

ΜΕΘΟΔΟΛΟΓΙΑ

Δεδομένου ότι όλες οι εφαρμοσθείσες μέθοδοι αξιολόγησης ήταν διαφορετικές από τη φύση τους και έδωσαν τόσο ποιοτικά όσο και ποσοτικά αποτελέσματα, ακολουθήθηκε μια «ελαχιστοποιητική» προσέγγιση για το συνδυασμό τους. Μόνο ποσοτικά αποτελέσματα επιλέχθηκαν για επεξεργασία και τα προβλήματα ευχρηστίας που αποκαλύφθηκαν κατηγοριοποιήθηκαν και ομαδοποιήθηκαν σε περισσότερες γενικές κατηγορίες προβλημάτων, έτσι ώστε να μπορούν να συγκριθούν και να αξιολογηθούν. Συνεπώς τελικά μελετήθηκε μόνο το ελάχιστο κοινό σύνολο προβλημάτων που όλες οι μέθοδοι μπόρεσαν να αποκαλύψουν. Μετά από αυτή την κατηγοριοποίηση, αναδύθηκε μια ομάδα από είκοσι ερωτήσεις, που ανήκουν στις πέντε κύριες κατηγορίες: *πλοήγηση και προσανατολισμός, διαχείριση στοιχείων της διασύνδεσης, πολυμεσικές πληροφορίες* (ποσότητα και ποιότητα), *βοήθεια και χρήση ασκήσεων*. Οι απόψεις των αξιολογητών απεικονίζονται για κάθε ερώτημα με τις μέσες τιμές (MV). Είναι γνωστό ότι η μέση τιμή είναι ευαίσθητη σε ακραίες τιμές (outliers), έτσι προτιμήθηκε αντί της διαμέσου (median), ώστε να μην αποκλειστεί η άποψη κανενός αξιολογητή. Μιας και ο κύριος σκοπός αυτής της εργασίας ήταν η συγκριτική αξιολόγηση των μεθόδων που χρησιμοποιήθηκαν, ο συντελεστής συσχέτισης Pearson- r χρησιμοποιήθηκε για τους υπολογισμούς μεταξύ των διαφόρων ομάδων δεδομένων.

Για να διευκρινιστεί η συσχέτιση μεταξύ των διαφορετικών προσεγγίσεων που εφαρμόστηκαν, έγιναν οι παρακάτω υποθέσεις.

Μηδενικές υποθέσεις:

H₀₁: Δεν υπάρχει καμία στατιστική συσχέτιση μεταξύ των δύο μεθοδολογιών (αξιολόγηση ειδικών-χρηστών)

H₀₂: Μόνο οι εμπειρικές μεθοδολογίες αξιολόγησης δίνουν στατιστικά σημαντική συσχέτιση.

Εναλλακτικές υποθέσεις:

H_{a1}: Υπάρχει στατιστική συσχέτιση μεταξύ των δύο μεθοδολογιών (αξιολόγηση ειδικών-χρηστών)

H_{a2}: Υπάρχει μία τουλάχιστον μέθοδος αξιολόγησης με ειδικούς που δίνει στατιστικά σημαντική συσχέτιση.

ΕΠΕΞΕΡΓΑΣΙΑ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

Θεωρήθηκαν αρχικά οι μέσες τιμές (MVexp and MVusab) των δύο μεθοδολογιών αξιολόγησης (αξιολόγησης ειδικών και εμπειρική αντίστοιχα). Οι τιμές τους βρίσκονται σε γραμμική σχέση, τα δείγματα της μελέτης συσχετίζονται (εφαρμόστηκαν στο ίδιο κομμάτι λογισμικού) και υπάρχει κλίμακα μέτρησης ίσων διαστημάτων. Άρα πληρούνται οι προϋποθέσεις και το κριτήριο Pearson-r είναι εφαρμόσιμο.

Ακολούθως υπολογίστηκε ο συντελεστής Pearson-r για τις ανωτέρω μέσες τιμές. Για 18 βαθμούς ελευθερίας, για υπόθεση διπλής κατεύθυνσης και για επίπεδο σημαντικότητας $p < .05$, το Pearson-r παρουσιάζει κρίσιμη τιμή 0,4438. Η επεξεργασία έδειξε τα εξής :

$$\text{MVexp} / \text{MVusab} \quad r_1 (18) = 0,60, p < .05 \quad (1)$$

Ακολούθως υπολογίστηκε η ολική μέση τιμή MV (MV gen) όλων των αξιολογήσεων. Ακολούθως, όλοι οι επόμενοι συσχετισμοί αναφέρονται σ' αυτή την μέση τιμή MV προκειμένου να ομογενοποιηθούν τα αποτελέσματα. Υποτίθεται εδώ ότι η τελική μέση τιμή όλων των αξιολογήσεων εκφράζει τον ακριβή αριθμό των πραγματικών προβλημάτων ευχρηστίας (για κάθε ερώτηση). Έτσι έχουν υπολογιστεί με όλους τους δυνατούς συνδυασμούς οι συντελεστές συσχέτισης Pearson-r. Παρακάτω παρουσιάζονται όμως μόνο εκείνοι που συμβάλλουν στις θετικές ερευνητικές υποθέσεις.

$$\text{MVexp} / \text{MVgen} \quad r_2 (18) = 0,79, p < .05 \quad (2)$$

$$\text{MVusab} / \text{MVgen} \quad r_3 (18) = 0,96, p < .05 \quad (3)$$

$$\text{CGW (Ελλάδα)} / \text{MVgen} \quad r_4 (18) = 0,56, p < .05 \quad (4)$$

$$\text{Phen (Ελλάδα)} / \text{MVgen} \quad r_5 (18) = 0,37, \text{ n.s.} \quad (5)$$

$$\text{TZlexp (Γερμανία)} / \text{MVgen} \quad r_6 (18) = 0,30, \text{ n.s.} \quad (6)$$

$$\text{QUIS (Ελλάδα)} / \text{MVgen} \quad r_7 (18) = 0,74, p < .05 \quad (7)$$

$$\text{TZlusab (Γερμανία)} / \text{MVgen} \quad r_8 (18) = 0,81, p < .05 \quad (8)$$

$$\text{Rota-obs (Ελλάδα)} / \text{MVgen} \quad r_9 (18) = 0,87, p < .05 \quad (9)$$

$$\text{Rota-QUIS (Ελλάδα)} / \text{MVgen} \quad r_{10} (18) = 0,84, p < .05 \quad (10)$$

$$\text{Holl-obs (Ολλανδία)} / \text{MVgen} \quad r_{11} (18) = 0,78, p < .05 \quad (11)$$

$$\text{Holl-que (Ολλανδία)} / \text{MVgen} \quad r_{12} (18) = 0,77, p < .05 \quad (12)$$

$$\text{Ergani-que (Ελλάδα)} / \text{MVgen} \quad r_{13} (18) = 0,76, p < .05 \quad (13)$$

Το πρώτο θέμα προς ερμηνεία είναι η συσχέτιση μεταξύ των αποτελεσμάτων των εμπειρικών και των με ειδικούς αξιολογήσεων. Μπορεί να υποστηριχθεί ότι δεν υπάρχει μεγάλη διαφορά ανάμεσα στην επίδοση των δύο προσεγγίσεων, αφού και οι δύο παρουσιάζουν στατιστικά σημαντική συνάφεια με το μέσο όρο, όπως δείχνουν οι ισότητες 2 και 3, αν και οι εμπειρικές μεθοδολογίες αξιολόγησης είναι πάντα προτιμητέες, αφού παρουσιάζουν σημαντική συνάφεια 0,96 (ισότητα 3). Έτσι το πρώτο αποτέλεσμα αυτής της μελέτης είναι ότι οι μέθοδοι αξιολόγησης ειδικών έχουν ωριμάσει αρκετά για να δώσουν μια αρκετά καλή εναλλακτική λύση σε σχέση με

τις εμπειρικές, ιδιαίτερα όσον αφορά τους πόρους που καταναλώνουν. Γενικά οι εμπειρικές προσεγγίσεις αξιολόγησης θεωρούνται περισσότερο αξιόπιστες και ποιοτικά πιο πολύτιμες, ωστόσο διαπιστώνονται συχνά προκαταλήψεις στις απόψεις των συμμετεχόντων, τις οποίες οι ειδικοί μπορούν να ξεπερνούν. Από την άλλη μεριά, οι ειδικοί τείνουν να υπερεκτιμούν ή να υποτιμούν την σοβαρότητα των προβλημάτων. Αυτό είναι φανερό στην περίπτωση της ανασκόπησης με ειδικούς (expert review) στην Γερμανία (ισότητα 6), που έδωσε ένα στατιστικά μη σημαντικό αποτέλεσμα εξ αιτίας της τυπικότητας με την οποία εκτελέστηκε και της αυστηρότητας των αξιολογητών. Η περίπτωση της φαινομενογραφικής (phenomenographical) προσέγγισης (ισότητα 5), η οποία επίσης έδωσε στατιστικά μη σημαντικά αποτελέσματα, βασίζεται, κατά την γνώμη μας, περισσότερο στην ποιοτική φύση της μεθόδου αυτής. Φαίνεται ότι είναι δύσκολα προσαρμόσιμη σε μια «συνδυαστική αξιολόγηση» με στατιστικό προσανατολισμό, μπορεί όμως να δώσει πολύτιμα ποιοτικά αποτελέσματα. Γενικά, όπως αναμενόταν, οι εμπειρικές μέθοδοι αξιολόγησης δίνουν έναν εγγύτερο συσχετισμό προς τη συνολική μέση τιμή, όπως δείχνουν οι ισότητες 8-13.

ΣΥΖΗΤΗΣΗ ΚΑΙ ΠΡΟΒΛΗΜΑΤΙΣΜΟΙ

Βασίζόμενοι στα προαναφερθέντα αποτελέσματα, πρέπει να απορρίψουμε τις μηδενικές υποθέσεις και να αποδεχθούμε τις εναλλακτικές. Υπάρχει μια στατιστικά σημαντική συσχέτιση μεταξύ εμπειρικών και αξιολογήσεων με ειδικούς στην περίπτωση της παρούσας έρευνας, η οποία οδηγεί στο συμπέρασμα ότι οι αξιολογήσεις με ειδικούς δίνουν ένα βιώσιμο υποκατάστατο σε σχέση με τις εμπειρικές αξιολογήσεις. Επιπρόσθετα, δύο από τις τέσσερις εφαρμοσθείσες μεθόδους αξιολόγησης με ειδικούς, απέδωσαν με στατιστικά ικανοποιητική συσχέτιση ως προς την συνολική μέση τιμή των ερωτήσεων που χρησιμοποιήθηκαν κατά την έρευνα (ισότητες 4 και 7). Γενικεύοντας αυτά τα συμπεράσματα, μπορεί να υποστηριχθεί ότι αποδεικνύεται και στατιστικά ότι οι εμπειρικές αξιολογήσεις έχουν υψηλότερη απόδοση. Ωστόσο μερικές προσεγγίσεις αξιολόγησης ειδικών θα μπορούσαν να δώσουν μια ικανοποιητική εναλλακτική λύση, ειδικά στα πρώτα στάδια του σχεδιαστικού κύκλου του λογισμικού, όπου είναι δύσκολο να εφαρμοσθούν προσεγγίσεις με χρήστες, όπως αναφέρει και η σχετική βιβλιογραφία (Nielsen, 1993; Shneiderman, 1998). Οι μεθοδολογίες αξιολόγησης ειδικών, όπως φαίνεται σ' αυτή την μελέτη, λειτουργούν ικανοποιητικά και καταναλώνουν πολύ λιγότερους πόρους. Όμως οι εμπειρικές προσεγγίσεις αξιολόγησης είναι αναπόφευκτες προκειμένου να εκμαιεύσουν πολύτιμες πληροφορίες σχετικές με την επίδοση του χρήστη. Τα αποτελέσματα αυτής της μελέτης δείχνουν ότι οι παρατηρητικές μέθοδοι (ισότητες 9 και 11) και τα τεστ ευχρηστίας (σε εργαστήριο ευχρηστίας, ισότητα 8) δίνουν την υψηλότερη απόδοση, ενώ οι μέθοδοι που βασίζονται σε ερωτηματολόγια (ισότητες 12 και 13) είναι λιγότερο ακριβείς. Όπως αναφέρεται στην βιβλιογραφία (πχ. Lewis & Rieman, 1994; Nielsen, 1993), αυτό μπορεί να οφείλεται στις προκαταλήψεις των συμμετεχόντων.

Τέλος, υπάρχουν επίσης κάποιοι προβληματισμοί που απορρέουν από αυτή τη μελέτη. Η προσέγγιση που ακολουθήθηκε στην αξιολόγηση των δεδομένων, ενέχει πολλούς περιορισμούς. Ο πιο σημαντικός είναι, το ότι λαμβάνει υπ' όψη μόνο τα ποσοτικά δεδομένα που συγκεντρώθηκαν κατά την διάρκεια της έρευνας σε μια προσπάθεια να συγκρίνει τις διαφορετικές μεθόδους αξιολόγησης. Ωστόσο πολύτιμα ποιοτικά δεδομένα έχουν παραλειφθεί, για παράδειγμα από τα ερωτηματολόγια και την φαινομενογραφική προσέγγιση. Επιπρόσθετα, θα μπορούσε να γίνει και περισσότερο εκτενής στατιστική επεξεργασία προκειμένου να ερευνηθούν άλλα θέματα, όπως τάσεις ή διαφορές μεταξύ των μεθοδολογιών. Τελικά, το αποτέλεσμα αυτής της μελέτης παρέχει μόνο μια πρώτη ένδειξη σχετικά με την συσχέτιση των μεθοδολογιών που χρησιμοποιήθηκαν. Ωστόσο, αυτή η συσχέτιση βασίζεται σε περισσότερες πηγές και σε μακροχρόνια αξιολόγηση, άρα μπορεί να θεωρηθεί σαν το πρώτο βήμα στην κατεύθυνση που πραγματεύεται.

ΣΥΜΠΕΡΑΣΜΑ

Συνοψίζουμε θεωρώντας ότι οι δύο προσεγγίσεις, εμπειρικές και ειδικών, φαίνεται να αποδίδουν συγκρίσιμα, αν και η εμπειρική μεθοδολογία είναι πάντα προτιμητέα. Το επίπεδο ωριμότητας του λογισμικού παίζει ζωτικό ρόλο στην επιλογή της προσέγγισης αξιολόγησης. Ωστόσο η ποικιλία των πηγών που χρειάζονται για να πραγματοποιηθεί ένα ολοκληρωμένο σέτ αξιολογήσεων, περιορίζει πρακτικά την δυνατότητα εφαρμογής ενός ευρέως φάσματος. Έτσι, αυτή η μελέτη προτείνει για τα αρχικά στάδια σχεδιασμού και κατασκευής μια Γνωστική Γραφική Περιδιάβαση (CGW) ακολουθούμενη στα μετέπειτα στάδια από ένα συνδυασμό παρατηρητικής αξιολόγησης και αξιολόγησης βασισμένης σε ερωτηματολόγια.

ΕΥΧΑΡΙΣΤΙΕΣ

Οι συγγραφείς επιθυμούν να ευχαριστήσουν όλους όσους συμμετείχαν από τις τρεις χώρες που εργάστηκαν στο πρόγραμμα «Ορέστης». Δεν υπάρχει αρκετός χώρος σ' αυτή την εργασία για να αναφερθούν όλα τα ονόματά τους, ωστόσο ένα ιδιαίτερο «ευχαριστώ» πρέπει να απευθυνθεί στην Νομική Σχολή του ΑΠΘ και την Ολλανδική Νομική Σχολή του University of Groningen, καθώς και στο Technologiezentrum Informatik (TZI) του Πανεπιστημίου της Βρέμης. Τέλος, (αλλά όχι λιγότερο σημαντικό), ένα μεγάλο «ευχαριστώ» σε όλους τους αξιολογητές που έλαβαν μέρος και στην ΕΕ που χρηματοδότησε αυτό το πρόγραμμα.

ΑΝΑΦΟΡΕΣ

- Aedo, I., Catenazzi, N. & Diaz, P. (1996). The evaluation of a hypermedia learning environment: The CESAR experience. *Journal of Educational Multimedia & Hypermedia*, 5(1), 49-72
- Benyon, D., Davies, G., Keller, L., and Rogers, Y. (1990). *A guide to usability-usability now!* Milton Keynes: The Open University.
- Karat, C., Campbell, R. & Fiegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation, 1992. *Proceedings of ACM CHI '92*. Monterey, CA, May 3-7, 397-404.
- Karoulis, A., and Pombortsis, A. (2000 November). Evaluating the Usability of Multimedia Educational Software for Use in the Classroom Using a «Combinatory Evaluation» Approach. *Proc. of Eden 4th Open Classroom Conference*, 20-21 Nov 2000, Barcelona, Spain.
- Karoulis, A., Demetriades, S., Pombortsis, A. (2000 February). The Cognitive Graphical Jogthrough – An Evaluation Method with Assessment Capabilities. *Applied Informatics 2000 Conference Proceedings*, 369-373. Innsbruck, Austria. Anaheim, CA: IASTED/ACTA.
- Lewis, C. and Rieman, J. (1994). *Task-centered User Interface Design - A practical introduction*, Retrieved in October 2000 from ftp.cs.colorado.edu/pub/cs/distrib/HCI-Design-Book.
- Nielsen, J. (1993). *Usability Engineering*. San Diego: Academic Press.
- Maier, N.R.F. (1931). Reasoning in humans: II. The solution of a problem and its appearance in consciousness. *Journal of Comparative Psychology*, 12, 181-194.
- Marton, F. (1988). Phenomenography: Exploring Different Conceptions of Reality. In D.M. Fetterman (Edt). *Qualitative Approaches to Evaluation in Education: The Silent Scientific Revolution*, New York: Praeger, 176-205.
- Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T., (1994). *Human-Computer Interaction*, Addison-Wesley Publishing Company.
- Report 1, 2000. *Internal report: «Orestis»: Bericht über die Praxisevaluation in den Jugendvollzugsanstalt*. Bremen Blockland, Universität Bremen, Institut für Kriminalpolitikforschung, Bremen.

Report 2, 2000. *Internal report: «Orestis»: Bericht über die Praxisevaluation in den Jugendvollzugsanstalten FOC Kolkemate und PI Zutphen*, University of Leiden, Holland.

Scriven, M. (1976). The methodology of evaluation. In Tyler, R. (edt.), *Perspectives of Curriculum Evaluation*, Rand McNally, Chicago.

Shneiderman, Ben. (1998). *Designing the User Interface*, 3rd ed., Reading, Mass: Addison-Wesley



