

Κεφάλαιο 1

Η Θεωρία και η Ταξινόμηση των Μεθοδολογιών Αξιολόγησης

1.1. Περίληψη

Στο Κεφάλαιο αυτό θα παρουσιάσουμε τη γενική θεωρία των αξιολογήσεων. Αρχικά θα δούμε τι είναι η αξιολόγηση, τις εκπαιδευτικές αξιολογήσεις και τις αξιολογήσεις των διασυνδέσεων με τις ιδιαιτερότητές τους. Στη συνέχεια θα αναφέρουμε τα χαρακτηριστικά των αξιολογήσεων, το πώς αυτά τα χαρακτηριστικά διαχωρίζουν τις αξιολογήσεις σε είδη, ενώ θα ασχοληθούμε αναλυτικότερα με το βασικό θέμα των σταδίων από τα οποία περνάει κάθε αξιολόγηση. Εν συνεχεία θα κάνουμε μια προσπάθεια ταξινόμησης των γνωστών μεθοδολογιών με παράθεση των πλεονεκτημάτων και μειονεκτημάτων της κάθε μιας ενώ τέλος θα ολοκληρώσουμε με μια συζήτηση για διάφορα γενικά θέματα σχετικά με τις αξιολογήσεις.

1.2. Τι είναι η αξιολόγηση (γενικά)

Σύμφωνα με τους Preece et al. (1994) η αξιολόγηση είναι γενικά μια διαδικασία συλλογής στοιχείων σχετικών με την ευχρηστία μιας σχεδίασης ή ενός προϊόντος για μια συγκεκριμένη ομάδα χρηστών και για μια ιδιαίτερη δραστηριότητα μέσα σε καθορισμένο περιβάλλον ή πλαίσιο εργασίας. Στον ορισμό αυτό δίνεται δηλαδή ιδιαίτερη βαρύτητα στο περιβάλλον εφαρμογής του συγκεκριμένου προϊόντος.

Σκοπός της αξιολόγησης είναι ο εντοπισμός και η πρόταση λύσεων για προβλήματα που δημιουργούνται από τη συγκεκριμένη σχεδίαση του προϊόντος (Δημητριάδης, 2000). Σαν πρόβλημα θεωρείται κάθε τί που παρεμβαίνει στη δυνατότητα του χρήστη να ολοκληρώνει εργασίες *αποτελεσματικά (effectively)* και *αποδοτικά (efficiently)* (Karat et al., 1992).

Επίσης, σύμφωνα με τον Μακράκη (1999) η αξιολόγηση είναι μια διαδικασία:

- συλλογής
- επεξεργασίας και ανάλυσης και
- ερμηνείας

δεδομένων

- που αφορούν συγκεκριμένα αντικείμενα
- με βάση συγκεκριμένα κριτήρια

με απώτερο σκοπό την ανατροφοδότηση των δραστηριοτήτων και προσπαθειών μας με χρήσιμες πληροφορίες. Καταλήγοντας, η αξιολόγηση μας επιτρέπει να αποφανθούμε αν η ίδια η δραστηριότητα ή η συγκεκριμένη υλοποίησή της αξίζει να επαναληφθεί, να απορριφθεί ή να τροποποιηθεί.

Αν και οι ορισμοί αυτοί είναι εύστοχοι και περιεκτικοί, όπως θα δούμε στη συνέχεια αυτού του κεφαλαίου, η διαδικασία της αξιολόγησης επεκτείνεται αρκετά περισσότερο σε πλάτος και βάθος. Για παράδειγμα το στάδιο της προετοιμασίας της αξιολόγησης είναι από τα ζωτικότερα και δεν μπορεί να παραληφθεί.

1.3. Οι εκπαιδευτικές αξιολογήσεις

Ειδικά η εκπαιδευτική αξιολόγηση είναι μια διαδικασία που αποσκοπεί στο να προσδιορίσει το βαθμό υλοποίησης των στόχων ενός εκπαιδευτικού συστήματος ή των συστατικών του μερών. (Κασσωτάκης, 1990, Chinapah & Miron, 1990, Δημητρόπουλος, 1991, Calder, 1995)

Εδώ πρέπει να γίνει μια σημαντική παρέμβαση. Η τοποθέτηση αυτή αναφέρεται σαφώς στην αξιολόγηση του περιεχομένου του προγράμματος σπουδών, πράγμα που φαίνεται να ενδιαφέρει ιδιαίτερα τους σχεδιαστές εκπαιδευτικών προγραμμάτων. Όμως, η διατριβή αυτή θέλει να προχωρήσει ένα βήμα παρακάτω. Έχοντας υπ' όψιν μας τις βασικές εκπαιδευτικές θεωρίες και ιδιαίτερα την τεχνολογική θεώρηση, όπου αναφέραμε ότι δεν αρκεί ένα τεχνολογικά «προχωρημένο» περιβάλλον για να φέρει αποτέλεσμα, αλλά χρειάζεται παράλληλα και μια συνεπής εκπαιδευτική θεώρηση (Duffy & Knuth, 1990; Spiro & Jeng, 1990; Bransford et al., 1990; Vosniadou, 1996), έτσι μπορεί κανείς να ισχυριστεί ότι επίσης δεν αρκεί ένα καλό περιεχόμενο για να δημιουργήσει ένα αποτελεσματικό εκπαιδευτικό περιβάλλον, αλλά αυτό πρέπει να συνοδεύεται και από ένα σωστό «κανάλι» μεταφοράς γνώσης, του οποίου ένα μεγάλο μέρος, ειδικά στην τεχνολογική θεώρηση της εκπαίδευσης, φέρει η διασύνδεση του χρήστη με το σύστημα, στην επεκταμένη της μορφή που παρουσιάζει αυτή η διατριβή.

Συνεπώς εδώ θα ασχοληθούμε επιφανειακά μόνο με την αξιολόγηση του συνολικού «εκπαιδευτικού πακέτου» (και κυρίως στο *Κεφάλαιο 6*) και θα επικεντρωθούμε ειδικότερα στην αξιολόγηση των εκπαιδευτικών διασυνδέσεων στα πλαίσια της τεχνολογικής θεώρησης της εκπαίδευσης. Όπως θα δούμε αργότερα, υπάρχουν πολλές μεθοδολογίες αξιολόγησης των διασυνδέσεων και πολλές από αυτές επεκτείνονται και στην αξιολόγηση του συνολικού εκπαιδευτικού πακέτου. Η αξιολόγηση που προτείνει για παράδειγμα ο Μακράκης (1999) και αναφέρθηκε στην αρχή είναι ένα είδος ευρετικής αξιολόγησης, βασισμένης σε μια λίστα κριτηρίων που θα κρίνουν ειδικοί, η οποία, επειδή είναι πολύ συνηθισμένη, και θα συζητηθεί επίσης σε αυτή τη διατριβή στο *Κεφάλαιο 6*.

1.4. Η αξιολόγηση των διασυνδέσεων (interface evaluation)

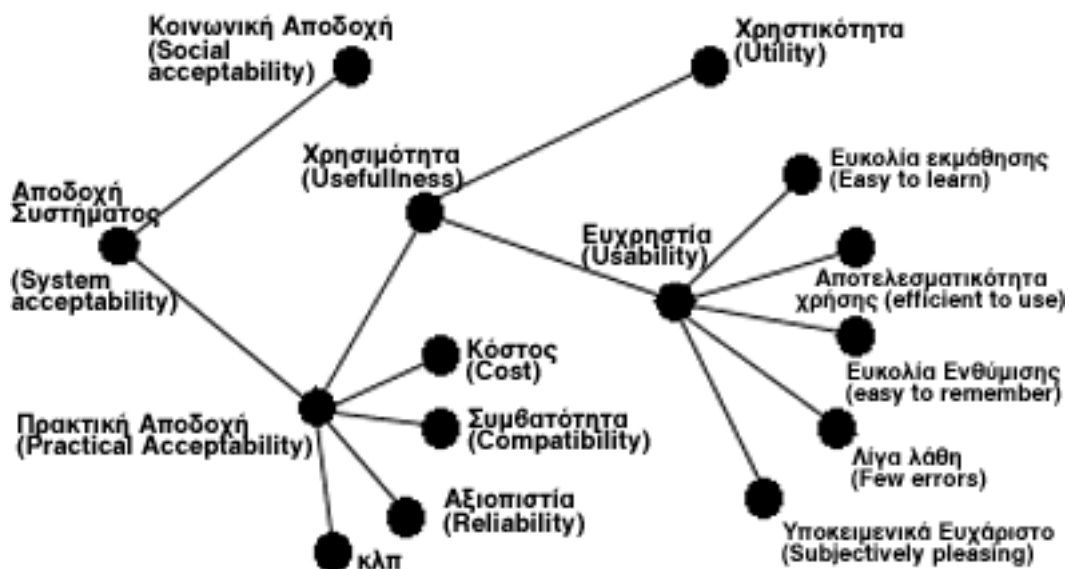
Όταν πρόκειται για αξιολόγηση που έχει σχέση με υπολογιστικά συστήματα και ιδίως όταν πρόκειται για αξιολόγηση διαδραστικών συστημάτων (interactive systems), τότε θεωρούμε ότι αφ' ενός ισχύουν όλες οι βασικές αρχές που αναφέρθηκαν, αφ' ετέρου όμως έχουμε επιπλέον και κάποιες πρόσθετες παραμέτρους που πρέπει να ληφθούν υπ' όψιν. Κατ' αρχάς, καθένας που χρησιμοποιεί ένα λογισμικό έχει σαν βασικό σκοπό, να λύσει ένα πρόβλημα, με άλλα λόγια να δουλέψει με το λογισμικό αυτό. Γι' αυτό το σκοπό πρέπει να χρησιμοποιήσει τη διασύνδεση του λογισμικού με το χρήστη, η οποία με τη σειρά της πρέπει να επιτρέψει στο χρήστη να ενεργοποιήσει όλο το δυναμικό του συστήματος (Benyon et al., 1990). Κάτω από αυτή τη θεώρηση συνεπώς, η διασύνδεση πρέπει:

- Να κάνει διαθέσιμα με έναν προφανή τρόπο όλα τα ισχυρά σημεία του λογισμικού (η αρχή της διαισθαντικότητας που ήδη αναφέρθηκε)
- Να είναι τελείως διάφανο στο χρήστη (αρχή της διαφάνειας). Δηλαδή ο χρήστης πρέπει να μπορεί να επικεντρώνεται στο τι θα κάνει και όχι στο πως θα το κάνει.

Με άλλα λόγια, σε κάθε σημείο η διασύνδεση πρέπει να βοηθάει και να παρέχει στο χρήστη, με προφανή τρόπο, την κατάλληλη ενέργεια για να πετύχει αυτός το στόχο του (Roth & Chair, 1997).

Για να μπορέσουμε συνεπώς να αποφανθούμε αν η διασύνδεση πληρεί αυτές τις προϋποθέσεις πρέπει να χρησιμοποιήσουμε κάποια προσέγγιση αξιολόγησης. Το κεφάλαιο αυτό στη συνέχεια ασχολείται με την παρουσίαση των βασικών μεθοδολογιών αξιολόγησης των διασυνδέσεων, τα βασικά βήματα που ακολουθούνται σε κάθε μεθοδολογία και την ταξινόμησή τους.

Σύμφωνα με τους Catenazzi et al. (1997) οι δύο βασικοί άξονες που ζητά να αποτιμήσει η αξιολόγηση είναι η ευχρηστία (usability) και η λειτουργικότητα (functionality). Οι έννοιες αυτές είναι αλληλένδετες καθώς στην ουσία η ευχρηστία θα καθορίσει πόσο καλά ο χρήστης θα μπορέσει να χρησιμοποιήσει τη λειτουργικότητα του συστήματος. Ειδικά, σχετικά με την ευχρηστία ο Nielsen (1993b) δίνει το επόμενο σχήμα:

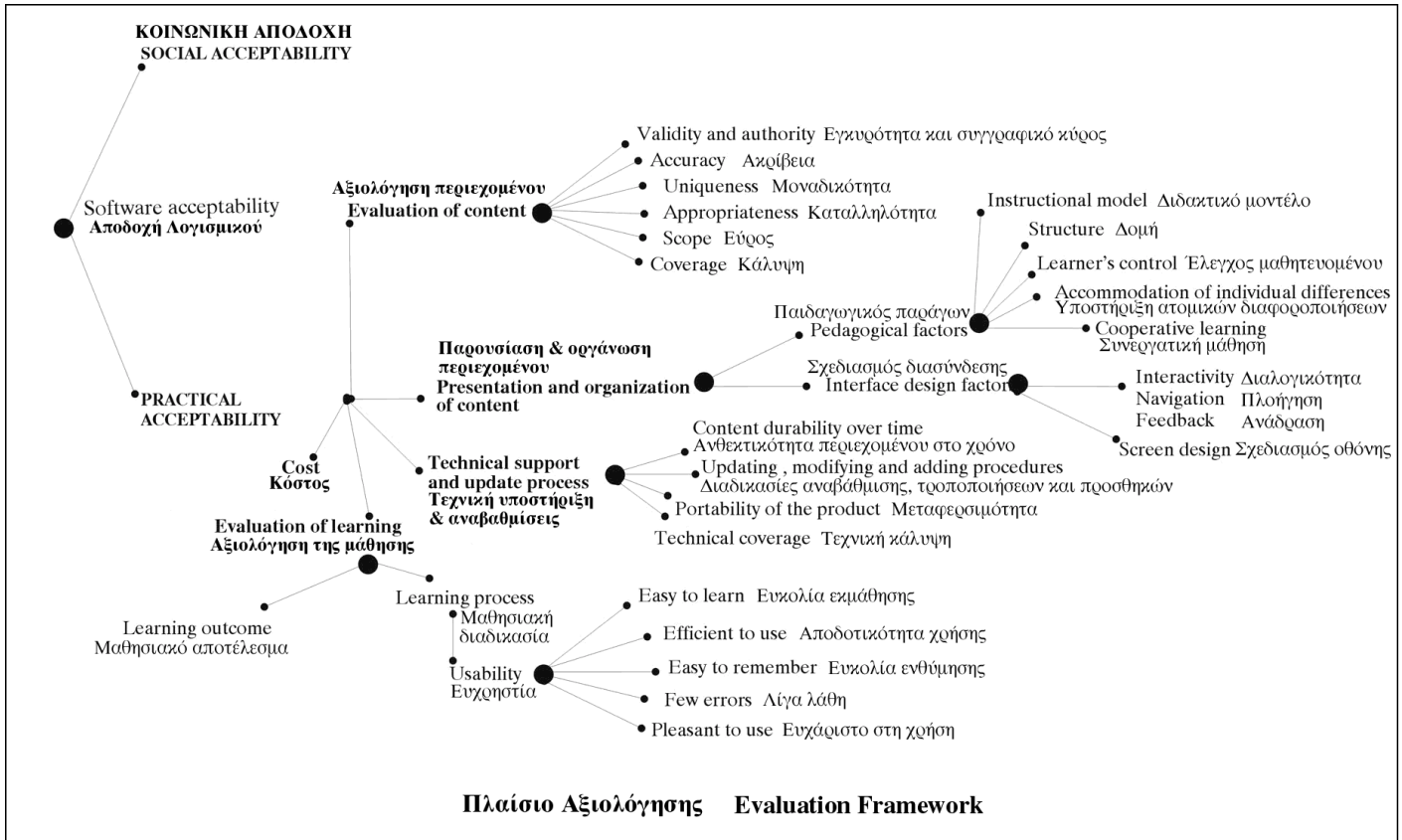


Σύμφωνα με αυτό το σχήμα, η ευχρηστία του λογισμικού είναι ένα μόνο μέρος της συνολικής αποδοχής του συστήματος (system acceptability), και για να επιτευχθεί πρέπει κανείς να ακολουθήσει τη διαδρομή (path) της πρακτικής αποδοχής του (practical acceptability) και της χρησιμότητάς του (usefulness) για να φτάσει στην ευχρηστία (usability) του συστήματος.

Η ευχρηστία με τη σειρά της αναλύεται στις εξής πέντε παραμέτρους:

- Την ευκολία εκμάθησης (easy to learn)
- Την αποδοτικότητα κατά τη χρήση του (efficient ot use)
- Την ευκολία να θυμάται κανείς τη χρήση του (easy to remember)
- Τον χαμηλό ρυθμό σφαλμάτων χρήστη (low error rate)
- Και την υποκειμενική ικανοποίηση χρήστη (subjective satisfaction)

Το σχήμα αυτό έχει επίσης προταθεί στην εκτεταμένη του μορφή από τους Georgiadou & Economides (2000) ώστε να καλύπτει και τα άλλα πεδία της αποδοχής του συστήματος – άρα τελικά να δίνει και ένα μέτρο της δυνητικής επιτυχίας του – ως εξής:



1.5. Βασικά χαρακτηριστικά των αξιολογήσεων

Αν κανείς θέλει να ταξινομήσει τα είδη των αξιολογήσεων, πρέπει πρώτα να διαχωρίσει τους κύριους άξονες σύμφωνα με τους οποίους κατηγοριοποιούνται αυτές.

Διαμορφωτικές και αποτιμητικές αξιολογήσεις

Αρχικά θεωρούμε την παράμετρο του χρονικού σημείου, κατά το οποίο θα λάβει χώρα η αξιολόγηση. Ο (Scriven, 1976) καθόρισε τις δύο μεγάλες κατηγορίες της διαμορφωτικής (formative) και συνολικής (ή τελικής ή αθροιστικής - summative) αξιολόγησης. Κατά αυτόν η διαμορφωτική είναι η κατάσταση όπου κάποιος παρατηρεί την εξέλιξη του συστήματος και εξετάζει τις αλλαγές που διαδραματίζονται, χρησιμοποιώντας τα δεδομένα για την εξέλιξη ή τη βελτίωση του συστήματος. Η αξιολόγηση αυτή εφαρμόζεται στη φάση κατασκευής του συστήματος και είναι μικρής κλίμακας και εντατική, με αξιολογητές ειδικούς ή χρήστες που αναφέρουν συνεχώς τα αποτελέσματα στους σχεδιαστές, οι οποίοι εφαρμόζουν συνήθως κάποια παραλλαγή επαναληπτικού σχεδιασμού (iterative design), όπως για παράδειγμα η σπιραλέτα που προτείνεται από αυτή τη διατριβή. Η αξιολόγηση αυτή δηλαδή εστιάζεται στο σχεδιασμό του συστήματος.

Αντιθέτως η *συνολική* (ή *αποτιμητική* ή *τελική*) διεξάγεται μετά την ολοκλήρωση του συστήματος, με σκοπό να προσδιοριστεί η έκταση της επιτυχίας των στόχων του. Αποτιμά την επιτυχία του συστήματος σε σχέση με τους τεθέντες στόχους, ερευνά τις συνθήκες που επιτυγχάνουν τα καλύτερα αποτελέσματα και δίνει μοντέλα χρήσης με την καλύτερη σχέση κόστους/απόδοσης. Η φάση αυτή είναι εκτεταμένη (σε χώρο και χρόνο), μεγάλης κλίμακας με περιστασιακές αναφορές και εστιάζεται στην εφαρμογή του συστήματος. (ΟΥ, 2000)

Οι (Chinapah & Miron, 1990) προτείνουν και τους όρους *αξιολόγηση διαδικασίας* (*process evaluation*) και *αξιολόγηση αποτελέσματος ή προϊόντος* (*product evaluation*), που στην ουσία είναι αντίστοιχοι με αυτούς του Scriven. Στην πρώτη περίπτωση δίνεται έμφαση στις διαδικασίες που συντελούνται προσπαθώντας να αναγνωρίσουμε τα προβλήματα και εμπόδια που παρεμβάλλονται στα στάδια σχεδιασμού και ανάπτυξης, έτσι ώστε να γίνουν οι απαραίτητες διορθωτικές παρεμβάσεις. Αντιθέτως, η αξιολόγηση αποτελέσματος εξετάζει αν υπάρχει αντιστοιχία μεταξύ αποτελέσματος και αρχικών στόχων και σε ποιο βαθμό.

Ποιοτικές και ποσοτικές αξιολογήσεις

Ο επόμενος διαχωρισμός που γίνεται στις αξιολογήσεις είναι η μορφή των αποτελεσμάτων που περιμένουμε να λάβουμε. Διακρίνουμε συνεπώς ανάμεσα σε *ποιοτικές* (*qualitative*) και *ποσοτικές* (*quantitative*) προσεγγίσεις. Στις ποιοτικές προσεγγίσεις το ενδιαφέρον επικεντρώνεται στην εξέταση των προσωπικών νοημάτων που αποδίδουν οι χρήστες σε αυτό που αξιολογείται. Στις ποσοτικές ζητάμε τη συλλογή ποσοτικά κατηγοριοποιημένων δεδομένων. Αυτά είναι δεδομένα στα οποία αποδίδεται μια αριθμητική τιμή ή αξία και στα οποία θα γίνει εφαρμογή στατιστικών μεθόδων και τεχνικών ανάλυσης (Μακράκης, 1999).

Υπάρχουν και στις δύο προσεγγίσεις πλεονεκτήματα και μειονεκτήματα και η συζήτηση γι' αυτά είναι έντονη και μακρά. Υπάρχουν υποστηρικτές και στις δύο όχθες. Υποστηρίζεται για παράδειγμα, ότι ο προκαθορισμός των θεωρητικών και μεθοδολογικών πλαισίων στην ποσοτική προσέγγιση περιορίζει την αξιολόγηση σε μια τεχνική διαδικασία πιστοποίησης ή απόρριψης του συνόλου ή μέρους των εννοιών, μεταβλητών ή παραγόντων που εικάζονται ότι συμβάλλουν στην εξήγηση ή πρόβλεψη του αξιολογουμένου φαινομένου. Υποστηρίζεται επίσης ότι η ποσοτική προσέγγιση αντιμετωπίζει τα φαινόμενα έξω από το φυσικό τους χώρο, χωρίς τη δυνατότητα διείσδυσης και ερμηνείας στα προσωπικά νοήματα που αποδίδουν οι χρήστες τους ή και οι ειδικοί σε αυτά (Μακράκης, 1999). Αντίστοιχα, οι ποιοτικές αξιολογήσεις είναι δυσκολότερα ερμηνεύσιμες και δεν υπόκεινται εύκολα σε στατιστική επεξεργασία. Επίσης, οι υποκειμενικές απόψεις των χρηστών υπόκεινται πολλές φορές σε προκαταλήψεις (*biases*) που μπορούν να διαστρεβλώσουν ακόμα και το αποτέλεσμα της έρευνας. Στο θέμα όμως αυτό θα επανέλθουμε και στο τέλος του κεφαλαίου αυτού, στη συζήτηση που θα ακολουθήσει.

Ανοικτές και κλειστές αξιολογήσεις

Ως προς το επικοινωνιακό κανάλι που θα εγκαταστήσουμε με τους εκτελούντες την αξιολόγηση διακρίνουμε γενικά δύο προσεγγίσεις, τις «*ανοικτές*», όπου ο χρήστης-αξιολογητής καλείται να εκφράσει τη γνώμη του χωρίς να ακολουθεί μια συγκεκριμένη δομή ερωτήσεων και τις «*κλειστές*» όπου ο χρήστης-αξιολογητής καλείται να ακολουθήσει μια προδιαγεγραμμένη πορεία ερωτήσεων. Οι προσεγγίσεις αυτές μπορούν να εφαρμοστούν σε όλες σχεδόν τις μεθοδολογίες που βασίζονται σε χρήστες, αφού οι αξιολογήσεις με ειδικούς συνήθως ακολουθούν προδιαγεγραμμένη πορεία. Επίσης είναι πιο συνηθισμένες στις ποιοτικές αξιολογήσεις και ειδικότερα κατά τη συλλογή των δεδομένων με ερωτηματολόγια, συνεντεύξη ή παρατήρηση. Κάποιες ιδιαιτερότητες αυτών των μεθόδων σχετικά με αυτούς τους τρόπους συλλογής δεδομένων σχολιάζονται παρακάτω.

Καταμετρητική και παρατηρητική αξιολόγηση

Για να καταγράψουμε τώρα τα αποτελέσματα της αξιολόγησης χρησιμοποιούμε πάλι δύο προσεγγίσεις, την *καταμετρητική (survey)* και την *παρατηρητική (observational)*.

Στην *καταμετρητική αξιολόγηση* καταγράφονται οι απόψεις του αξιολογητού, είτε από τον ίδιο σε ένα ερωτηματολόγιο, είτε από τους διεξάγοντες την αξιολόγηση μέσω συνήθως μιας συνέντευξης.

- ♦ Στο *ερωτηματολόγιο* οι κλειστές ερωτήσεις παρουσιάζουν τον κίνδυνο να κατευθύνουν την απάντηση σύμφωνα με τις προσδοκίες αυτών που διεξάγουν την έρευνα-αξιολόγηση (bias). Από την άλλη μεριά, αν και η ποιότητα των δεδομένων που συλλέγονται με τη μέθοδο του ανοικτού τύπου ερωτηματολογίου είναι συνήθως υψηλότερη από αυτή των κλειστού τύπου ερωτηματολογίων, οι ανοικτού τύπου ερωτήσεις πολλές φορές δημιουργούν προβλήματα κατηγοριοποίησης, ανάλυσης και ερμηνείας.
- ♦ Στη *συνέντευξη* βασικός είναι ο βαθμός εμπάθυνσης. Οι συνεντευξείς μπορούν να αναφέρονται σε “*δομημένες ή κλειστές*” ερωτήσεις, όπου ο ερωτώμενος απαντά συγκεκριμένα, χωρίς τη δυνατότητα εμπάθυνσης στο περιεχόμενο της απάντησης. Στις περιπτώσεις που ο ερευνητής-αξιολογητής στοχεύει στην εμπάθυνση, τότε οι “*ημι-δομημένες*” και οι “*ανοικτές*” συνεντεύξεις είναι περισσότερο κατάλληλες.

Στις *παρατηρητικές αξιολογήσεις* διακρίνουμε επίσης δύο διαφορετικές προσεγγίσεις.

- ♦ *Παρατήρηση χρηστών.* Χρήστες-αξιολογητές χρησιμοποιούν την υπό συζήτηση οντότητα ενώ παράλληλα ανεξάρτητοι παρατηρητές (συνήθως οι διεξάγοντες την αξιολόγηση) ή/και λογισμικό υπολογιστή ή άλλη καταγραφική μέθοδος καταγράφουν τα συμβαίνοντα και τις αντιδράσεις των χρηστών-αξιολογητών. Οι τρόποι και τα εργαλεία καταγραφής ποικίλουν και εξαρτώνται ισχυρά από την υπό αξιολόγηση οντότητα. Μπορούμε να έχουμε *δομημένη παρατήρηση*, όπου οι χρήστες-αξιολογητές ακολουθούν μια προδιαγεγραμμένη εργασιακή πορεία με την οντότητα, με σκοπό να εξαχθούν συγκεκριμένα συμπεράσματα, *ημι-δομημένη παρατήρηση*, όπου οι χρήστες-αξιολογητές προσπαθούν να πετύχουν εργαζόμενοι ένα τεθέντα στόχο, με ανεξάρτητο τρόπο όμως ο καθένας, σύμφωνα με την κρίση του, και *ελεύθερη παρατήρηση*, όπου οι χρήστες-αξιολογητές εργάζονται ελεύθερα με την οντότητα, πολλές φορές σε πραγματικές συνθήκες εργασίας.
- ♦ *Παρατήρηση ειδικών.* Ειδικοί στον υπό αξιολόγηση τομέα παρατηρούν και μελετούν την υπό αξιολόγηση οντότητα και καταγράφουν τις απόψεις τους. Ο βαθμός ελευθερίας που παρέχεται στον ερευνητή-αξιολογητή για το τι θα παρατηρήσει και πώς θα το καταγράψει ποικίλλει ανάλογα με τη φύση του αντικειμένου, τους στόχους και τους άξονες της αξιολόγησης. Η διαδικασία της παρατήρησης σε ορισμένες περιπτώσεις μπορεί πάλι να είναι *ημι-δομημένη*, με την έννοια ότι έχουμε μεγαλύτερα περιθώρια επιλογής στην παρατήρηση απ’ ό,τι στη *δομημένη παρατήρηση*, όπου η προκαθορισμένη διαδικασία δεν αφήνει περιθώρια επιλογής των παρατηρήσεων. Στη *μη δομημένη ή ελεύθερη παρατήρηση*, ο ερευνητής-αξιολογητής παρατηρεί το αντικείμενο αξιολόγησης κατά την πλήρη χρήση του για κάποιο χρονικό διάστημα και καταγράφει τα αποτελέσματα των παρατηρήσεών του.

1.6. Τα είδη των αξιολογήσεων

Αφού πλέον διευκρινίστηκαν οι βασικοί άξονες πάνω στους οποίους κινούνται όλα τα είδη των αξιολογήσεων, μπορούμε να ξεκινήσουμε μια προσπάθεια ταξινόμησης των αξιολογήσεων.

Δεν είναι όλες οι μεθοδολογίες αξιολόγησης της ίδιας μορφής, ούτε εφαρμόζονται στις ίδιες συνθήκες. Σύμφωνα με την κατηγοριοποίηση των Benyon et al. (1990) υπάρχουν πέντε γενικές κατηγορίες αξιολόγησης:

- **Αναλυτική αξιολόγηση (analytic evaluation).** Με μια περιγραφή της διασύνδεσης ή ένα πρωτότυπο ανά χείρας προχωρούμε σε εκτίμηση της επίδοσης του χρήστη κατά το χειρισμό της διασύνδεσης.
- **Αξιολόγηση των ειδικών (ή με ειδικούς) (expert evaluation).** Ειδικοί καλούνται να κρίνουν τη διασύνδεση παίρνοντας το ρόλο υποψηφίων χρηστών και να ανακαλύψουν τα προβλήματα που θα αντιμετωπίσουν αυτοί κατά το χειρισμό της.
- **Εμπειρική παρατηρητική αξιολόγηση (empirical observational evaluation).** Χρησιμοποιούνται χρήστες που δουλεύουν με τη διασύνδεση. Κατά τη διάρκεια της δουλειάς τους παρατηρούνται και οι αντιδράσεις τους καταγράφονται, είτε από παρατηρητές, είτε από συσκευές.
- **Εμπειρική καταμετρητική αξιολόγηση (empirical survey evaluation).** Χρησιμοποιούνται χρήστες που δουλεύουν με τη διασύνδεση. Στη συνέχεια ή ταυτόχρονα εκφράζουν τη γνώμη τους για θέματα χειρισμού η οποία καταγράφεται.
- **Πειραματική αξιολόγηση (experimental evaluation).** Ο αξιολογητής-ερευνητής μεταβάλλει παραμέτρους του συστήματος και βλέπει την επίδρασή τους στην απόδοση των χρηστών.

1.7. Στάδια της Αξιολόγησης

Η αξιολόγηση, με οποιαδήποτε μεθοδολογία και να επιλέξει κανείς, είναι μια διαδικασία που αποτελείται από αρκετά διαδοχικά στάδια, κατά τη διάσχιση των οποίων απαιτείται η ρύθμιση αρκετών παραμέτρων για να είναι αποτελεσματική. Τα βασικά αυτά στάδια είναι κατά τους Aedo et al. (1996):

- Η επιλογή μεθόδου
- Η προετοιμασία της αξιολόγησης
- Η διεξαγωγή της και
- Η αποτίμηση των αποτελεσμάτων της.

Κάθε ένα όμως από αυτά τα στάδια αναλύεται με τη σειρά του σε επί πλέον ενέργειες που το ολοκληρώνουν. Έχουμε έτσι για την πλήρη διεξαγωγή της αξιολόγησης μια σειρά ενεργειών όπως η ακόλουθη (ΟΥ, 2000):

1. Καθορισμός του στόχου, αξόνων και κριτηρίων της αξιολόγησης.

Όπως προείπαμε, στην παρούσα διατριβή μελετάμε πολυμεσικές, εκπαιδευτικές και διαδικτυακές διασυνδέσεις ανθρώπου-υπολογιστή ως προς την ευχρηστία τους, συνεπώς συνήθης στόχος όλων των αξιολογήσεων που μελετάμε εδώ είναι η ευχρηστία του συστήματος. Πάντως ο στόχος της αξιολόγησης θα πρέπει να είναι ο ίδιος με τις τελικές απαντήσεις που ζητάμε να πάρουμε στην αποτίμηση των αποτελεσμάτων. Συνεπώς εδώ χρειάζεται καλός καθορισμός του τι είναι και τι δεν είναι εφικτό να γίνει, όπως και το να διατηρούνται αυτές οι προδιαγραφές καθ' όλη τη διάρκεια της αξιολόγησης στο μυαλό του αξιολογητή-ερευνητή.

Οι άξονες τώρα είναι προφανώς οι συνιστώσες στις οποίες αναλύεται η ευχρηστία (ή ο όποιος άλλος στόχος έχουμε θέσει), και οι οποίες είναι, όπως προείπαμε, η ευκολία εκμάθησης, η αποδοτικότητα χρήσης, η ευκολία να θυμάται κανείς τη χρήση του, ο χαμηλός

ρυθμός σφαλμάτων χρήστη και τέλος η υποκειμενική ικανοποίηση χρήστη (Nielsen, 1990; ISO, 1998).

Τέλος τα *κριτήρια* σύμφωνα με τα οποία αξιολογούμε είναι ακριβώς οι εφαρμοσμένες διεργασίες που θα εκτελέσει ο χρήστης, αλλά *στην ιδανική τους μορφή*. Θεωρούμε δηλαδή ότι υπάρχει ο τέλειος τρόπος χρήσης της οντότητας, σύμφωνα με όλους τους κανόνες και τις προδιαγραφές σχεδιασμού και θα προσπαθήσουμε να συγκρίνουμε με αυτόν την πραγματική μορφή που λαμβάνουν οι διεργασίες μέσα στη συγκεκριμένη υλοποίηση της υπό συζήτηση οντότητας. Τα κριτήρια αυτά μπορεί συνεπώς να είναι είτε κανόνες προς τους οποίους οφείλει να συμμορφώνεται η οντότητα, είτε πρότυπες διεργασίες, αποδεδειγμένα εύχρηστες, όπως για παράδειγμα η *άμεση διαχείριση (direct manipulation)* (Shneiderman, 1998) για τη διαχείριση των εικονιδίων με το ποντίκι.

Υπάρχουν κάποιες παράμετροι που μπορούν να βοηθήσουν για να οριστούν οι στόχοι, οι άξονες και τα κριτήρια, όπως οι στόχοι που έχει η αρχική εφαρμογή, οι αναμενόμενοι χρήστες και οι ανάγκες τους, ακόμα και το να ορίσουμε κάποιες «κλάσεις» χρηστών.

Το πόσους στόχους θα θέσει κανείς είναι ένα επίσης σημαντικό ζήτημα. Το να θέσει κανείς πυκνότητα στόχων σε επίπεδο ενεργειών μπορεί να οδηγήσει σε προβλήματα (Wharton et al., 1992). Είναι καλύτερο να οριστούν οι στόχοι σε επίπεδο διεργασιών, ή και σε ανώτερο επίπεδο, για παράδειγμα μια σειρά διεργασιών που οδηγούν σε κάποιο ολοκληρωμένο στόχο της εφαρμογής. Το θέμα αυτό θα συζητηθεί αναλυτικότερα στο *Κεφάλαιο 2*.

2. Επιλογή της μεθόδου αξιολόγησης.

Όπως έχει ήδη καταστεί προφανές, δεν συνιστάται κάθε μέθοδος για κάθε περίπτωση. Παράγοντες που επηρεάζουν την επιλογή μεθόδου είναι (Aedo et al., 1996):

1) *Το στάδιο ανάπτυξης του interface*. Στα αρχικά στάδια ανάπτυξης ή και στον σχεδιασμό ακόμα του περιβάλλοντος διασύνδεσης, προέχει να αξιολογηθούν οι βασικές διεργασίες (tasks) και η χρησιμότητά τους. Αντιθέτως στα πλέον προχωρημένα στάδια αξιολογείται το πόσο προφανείς είναι αυτές οι διεργασίες, πόσο αποδοτικές είναι στην κατεύθυνση της επίτευξης του τελικού στόχου του χρήστη, δηλαδή η γενικότερη ευχρηστία τους. Στα αρχικά στάδια συνεπώς προτιμητέες είναι οι αναλυτικές μέθοδοι, στα μέσα στάδια σχεδιασμού οι αξιολογήσεις με ειδικούς και οι εμπειρικές και στα τελικά οι με ειδικούς οι εμπειρικές και οι πειραματικές μέθοδοι.

2) *Πλήθος και τύπος χρηστών*. Αν μας ενδιαφέρει η επίδοση των χρηστών ή αν έχουμε μεγάλους αριθμούς χρηστών, χρησιμοποιούμε αναλυτικές μεθόδους για εξοικονόμηση πόρων κατά την μετέπειτα χρήση του συστήματος. Αν πάλι η εφαρμογή μας απευθύνεται σε περιορισμένο αριθμό χρηστών ή είναι ειδικής μορφής, τότε μόνο αξιολόγηση με πραγματικούς χρήστες θα δώσει αξιόπιστα αποτελέσματα.

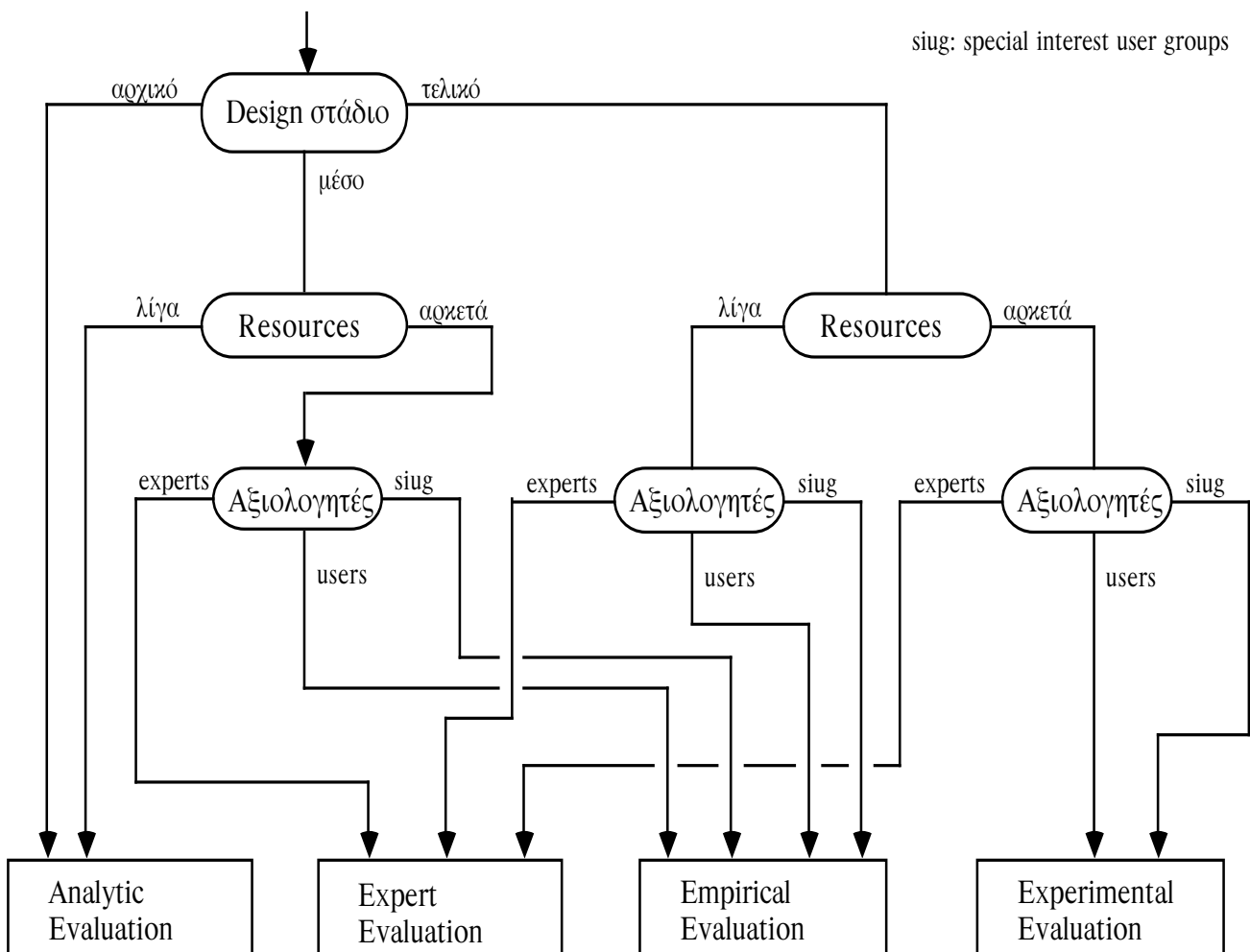
3) *Οι κατάλληλοι ή διαθέσιμοι αξιολογητές*. Αφού πρώτα καθοριστεί ο πληθυσμός και το είδος χρηστών που απευθυνόμαστε (target user group) αποφασίζεται το είδος των απαραίτητων αξιολογητών. Συνήθως απευθυνόμαστε σε ένα από τα εξής τρία είδη αξιολογητών: ειδικούς (experts), κοινούς χρήστες και ειδικούς χρήστες. Οι ειδικοί μπορεί να είναι ή να μην είναι μέλη της ομάδας σχεδιασμού του συστήματος, όμως σε κάθε περίπτωση συνιστάται να είναι άνθρωποι με εμπειρία στο HCI ή στο πεδίο εφαρμογής της υπό αξιολόγηση διασύνδεσης, ακόμα καλύτερα και στα δύο. Συνιστάται επίσης, όπως θα δούμε και αργότερα, η δόμηση μιας διεπιστημονικής ομάδας αξιολογητών ή οποία θα είναι σε θέση να δει το σύστημα από διαφορετικές οπτικές γωνίες. Οι απλοί χρήστες είναι οι χρήστες γενικού λογισμικού, ενώ με την έννοια ειδικοί χρήστες εννοούμε χρήστες που θα χρησιμοποιήσουν κάποιο λογισμικό ειδικής χρήσης, πχ. έλεγχος εναέριας κυκλοφορίας, ή διαχείριση χαρτοφυλακίου πελατών. Όμως, όπως πλέον συμβαίνει σχεδόν παντού, τα όρια

ανάμεσα στις διάφορες κατηγορίες τείνουν προς το ασαφές. Για παράδειγμα, οι μαθητές που θα χρησιμοποιήσουν τα λογισμικά του σχολείου τους είναι γενικοί ή ειδικοί χρήστες;

4) Το είδος των προσδοκωμένων αποτελεσμάτων. Αν το ζητούμενο είναι η εξοικονόμηση λειτουργικών κόστους, τότε εφαρμόζουμε αναλυτικές μεθόδους, αν μας ενδιαφέρει η ακρίβεια ή πρόκειται για ζωτικής σημασίας (life-critical) εφαρμογή, τότε η αξιολόγηση με χρήστες είναι ίσως η καταλληλότερη, αν πάλι ζητάμε την αποκάλυψη μόνο των σοβαρών προβλημάτων, η αξιολόγηση με ειδικούς είναι η συνιστώμενη μέθοδος.

5) Πόροι και εξωτερικοί περιορισμοί, όπως διαθέσιμος χρόνος για αξιολόγηση, διαθέσιμο κόστος και διαθέσιμος εξοπλισμός. Σε περίπτωση στενότητας πόρων, αδιαφιλονίκητος νικητής είναι οι μέθοδοι βασισμένες σε ειδικούς, αφού συντομότερες χρονικά μέθοδοι είναι οι με ειδικούς, οικονομικότερες πάλι οι με ειδικούς και τους λιγότερους πόρους γενικά χρειάζονται επίσης οι αναλυτικές και οι με ειδικούς.

Παρακάτω δίνεται ένα πρόχειρο και ενδεικτικό διάγραμμα που ενοποιεί όλες τις αναφερθείσες παραμέτρους για την επιλογή της μεθόδου αξιολόγησης:



Παρατηρήσεις επί του διαγράμματος:

Στάδιο σχεδιασμού. Σε ποιο στάδιο σχεδιασμού βρίσκεται το σύστημα; Αυτό βέβαια εξαρτάται ισχυρά από τη μεθοδολογία σχεδιασμού που ακολουθείται, όπως για παράδειγμα η σπιραλέτα ή το star life cycle. Γενικά όμως μπορούμε να διακρίνουμε τις εξής φάσεις:

- Καθορισμός προδιαγραφών συστήματος (system specification)
- Σχεδιασμός του συστήματος.
- Πρώτο πρωτότυπο (first prototype)

- Δεύτερη έκδοση (beta version)
- Τελικό προϊόν (final product)

Αξιολογητές.

- Ειδικοί
- Χρήστες. Είναι οι “απλοί” (γενικοί) χρήστες, οι τελικοί καταναλωτές.
- Siug (special interest user groups): ειδικοί χρήστες (άτομα με ειδικές ανάγκες, επαγγελματίες κλπ)

Πόροι. Όλοι οι διαθέσιμοι πόροι, καθώς και οι ενυπάρχοντες περιορισμοί (χρονικοί, υλικοί, οικονομικοί κλπ)

- Υπάρχουν διαθέσιμοι κατάλληλοι αξιολογητές;
- Πόσος χρόνος υπάρχει;
- Υπάρχει ο απαραίτητος εξοπλισμός;
- Υπάρχει οικονομική ευχέρεια και σε τι ύψος;
- Εμπειρία αξιολογητών; (θεωρητικές γνώσεις πχ. στην πειραματική μέθοδο)

Τέλος, ακόμα και αν καταλήξει κανείς στη μεθοδολογία που θα εφαρμόσει, παραμένει το πρόβλημα της επιλογής της ακριβούς μορφής που πρέπει να εφαρμοστεί. Για παράδειγμα, στις αξιολογήσεις με χρήστες ακολουθούμε αρκετές διαφορετικές προσεγγίσεις για την καταγραφή των γενομένων: λογισμικό καταγραφής (logging software), ερωτηματολόγια, συνεντεύξεις, σημειώσεις παρατηρητών, σημειώσεις αξιολογητών και άλλες. Το ερώτημα που εύλογα γεννάται είναι ποιιά προσέγγιση να ακολουθηθεί κανείς. Εδώ υπάρχουν δύο ερωτήματα που μας διευκολύνουν να αποφασίσουμε:

- Το μέγεθος της ανάγκης να καταγραφεί **τι** κάνει ο χρήστης και
- Το μέγεθος της ανάγκης να καταγραφεί το **γιατί** το κάνει ο χρήστης.

Κάθε εργαλείο καταγραφής μπορεί να συμβάλλει σε διαφορετικό βαθμό προς αυτή την κατεύθυνση. Για παράδειγμα, το λογισμικό καταγραφής μπορεί να μας εφοδιάσει με μεγάλες ποσότητες δεδομένων σχετικά με τις ενέργειες του χρήστη, να απαντήσει δηλαδή στο **τι** έκαναν οι χρήστες, όμως δεν μπορεί να απαντήσει στο **γιατί** οι χρήστες τα έκαναν όλα αυτά.

Η ταξινόμηση των αξιολογήσεων που παρατίθεται στην παράγραφο 1.8. είναι το επόμενο σημείο που πρέπει να συμβουλευτεί κανείς για να αποφασίσει για την τελική μέθοδο, αφού έχει πλέον μια ιδέα για το είδος της αξιολόγησης που θέλει να εφαρμόσει. Στο σημείο εκείνο παρατίθεται μια σύντομη περιγραφή της κάθε προσέγγισης, τα συγκριτικά πλεονεκτήματα και μειονεκτήματα της κάθε μεθόδου, όπως αναφέρονται στη βιβλιογραφία, καθώς και οι περιπτώσεις που συνιστάται η κάθε μία από τις αναφερόμενες μεθόδους αξιολόγησης.

3. Προετοιμασία της αξιολόγησης

Η προετοιμασία της αξιολόγησης είναι τόσο βασικό στάδιο, ώστε πολλές φορές μπορεί να επηρεάσει έντονα τα αποτελέσματα. Είναι συνεπώς απαραίτητο να δίνεται η πρέπουσα βαρύτητα σ’ αυτό το στάδιο της αξιολόγησης.

α. Επιλέγουμε τους αξιολογητές (ειδικούς, χρήστες ή ειδικούς χρήστες)

Προφανώς η ομάδα των αξιολογητών εξαρτάται απόλυτα από τη μεθοδολογία αξιολόγησης που έχει επιλεγεί για να εφαρμοστεί. Στην περίπτωση εμπειρικής αξιολόγησης με χρήστες, το πρώτο στάδιο είναι να καταγραφεί η κατάλληλη *διαμόρφωση χρήστη (user profile)* και στη συνέχεια να ανευρεθούν οι ανάλογοι χρήστες που πληρούν αυτές τις προϋποθέσεις και να

εξασφαλιστεί η συμμετοχή τους. Το στάδιο αυτό αποδείχθηκε στην πράξη από τα δυσκολότερα και αποτελεί ένα από τα μειονεκτήματα των εμπειρικών αξιολογήσεων. Όμως, όπως ισχυρίζονται οι Lewis & Rieman (1994), «αν δεν μπορείτε να βρείτε αξιολογητές-χρήστες για το σύστημά σας, τι σας κάνει να πιστεύετε ότι θα βρείτε αργότερα πραγματικούς χρήστες για να το χρησιμοποιήσουν;» Καθορίζουμε συνεπώς πρώτα σωστά το προφίλ των χρηστών μας και στη συνέχεια απευθυνόμαστε στις ομάδες που προέκυψαν για την ανεύρεση των πραγματικών αξιολογητών, προσπαθώντας να συλλέξουμε όσο το δυνατόν πιο αντιπροσωπευτικό δείγμα χρηστών για την αξιολόγησή μας (Catenazzi et al., 1997)

Στις αξιολογήσεις με ειδικούς συνιστάται, όπως προαναφέρθηκε, η εμπλοκή ενός ή περισσότερων αξιολογητών με γνώσεις στο HCI ή στη γνωστική επιστήμη. Επίσης ο αριθμός των αξιολογητών σ' αυτού του είδους τις αξιολογήσεις φαίνεται να συγκλίνει στους τέσσερις με έξι αξιολογητές, διότι αυτό το μέγεθος ομάδας βοήθησε να διατηρηθεί η σωστή εστίαση κατά τη διάρκεια της διεξαγωγής της συνεδρίας (Wharton et al, 1992; Karoulis et al., 2000; Karoulis et al., 2000b)

β. Επιλέγουμε τις διεργασίες που θα αξιολογηθούν.

Το θέμα αυτό είναι αρκετά δύσκολο και περίπλοκο και δεν υπάρχουν σαφείς οδηγίες για τον τρόπο επιλογής σχεδόν από καμμία μέθοδο. Οι Jeffries et al. (1991) αντιμετώπισαν σοβαρό πρόβλημα στο σημείο αυτό και προτείνουν η έρευνα να στραφεί προς την ανεύρεση μεθοδολογίας ικανής να προτείνει τις καταλληλότερες διεργασίες και ενέργειες προς αξιολόγηση. Στην πράξη όμως αυτό αποδεικνύεται πολύ δύσκολο, αφού εξαρτάται απόλυτα από την υπό αξιολόγηση οντότητα. Κεντρικό σημείο είναι το ποιές ακριβώς ενέργειες κατά τη χρήση της οντότητας είναι αυτές που μας ενδιαφέρει να αξιολογήσουμε. Αναφερθήκαμε παραπάνω στους στόχους, τους άξονες και τα κριτήρια της αξιολόγησης. Στο σημείο αυτό μας ενδιαφέρει να ορίσουμε τρόπους χρήσης της οντότητας που να μπορούν να αποτιμηθούν με βάση τα τεθέντα κριτήρια. Ίσως μια προσέγγιση είναι η απάντηση κάποιων σαφών ερωτήσεων, όπως (Wharton et al., 1992):

- *Σε τί "ύψος" θα εκτελέσουμε την αξιολόγηση;* Θα δούμε την οντότητα σαν όλον, θα την διακρίνουμε σε συστατικά ή θα την αξιολογήσουμε σε επίπεδο διεργασιών που εκτελεί ο χρήστης. Η πιο συνηθισμένη περίπτωση είναι η τελευταία, και αυτό υποθέτουμε σε αυτή τη διατριβή, όμως πρέπει κανείς να αναρωτηθεί αν αυτή πράγματι είναι η πιο κατάλληλη προσέγγιση για την αξιολόγηση της υπό συζήτηση οντότητας.
- *Πόσο ρεαλιστικές και περίπλοκες πρέπει να είναι οι διεργασίες προς αξιολόγηση;* Γενικά συνιστάται να αξιολογούμε και απλές, αλλά και σύνθετες διεργασίες. Μια τεχνική που βοηθάει εδώ είναι να αξιολογούνται αρχικά απλές διαδικασίες που αποτελούν ομάδα, οπότε στη συνέχεια, κάθε φορά που θα ξανασυναντιώνονται, θα αξιολογείται η ομάδα ολόκληρη σαν ολότητα, επιταχύνοντας έτσι τη διεργασία (Rowley & Rhoades, 1992; Karoulis et al., 2000)
- *Πού πρέπει να μπουν τα όρια;* Πρέπει δηλαδή να αξιολογήσουμε το πλήρες δυναμικό του συστήματος ή μόνο το βασικό του πυρήνα; Αξιολογούμε τι μπορεί να κάνει το σύστημα ή τι περιμένει ο χρήστης ότι θα κάνει;
- *Πόσες διεργασίες είναι αρκετές;* Αν θέλουμε να καλύψουμε όλο το φάσμα των διεργασιών που μπορεί να εκτελέσει το σύστημα, μπορεί να χρειαστούμε 100 ώρες αξιολογήσεων. Από την άλλη, περισσότερο συρρικνωμένο σετ διεργασιών απ' ότι πρέπει θα οδηγήσει ενδεχομένως σε αναξιόπιστα ή μη γενικεύσιμα συμπεράσματα.
- *Τι θα κάνουμε με τις εναλλακτικές περιπτώσεις;* Γενικά δεν μπορούμε να υποθέσουμε ότι ο κάθε χρήστης θα ακολουθήσει οπωσδήποτε την προκαθορισμένη πορεία. Μπορούμε βέβαια να επιλέξουμε την πιθανότερη διαδρομή ή τρόπο που θα ακολουθήσει κατά την εκτίμησή μας ο μέσος χρήστης (ό,τι και να σημαίνει αυτό), αλλά πάντα θα υπάρχουν διαδρομές που μπορούν να παρουσιάσουν προβλήματα.

- *Ποιά είναι η σωστή πυκνότητα αξιολόγησης για κάθε διεργασία;* Όταν ζητείται για παράδειγμα από το χρήστη να εισάγει ένα κείμενο, αξιολογούμε την ταχύτητα εισαγωγής της λέξης ή κάθε γράμματος ξεχωριστά; Τι θα ορίσουμε σαν “ενέργεια” σε μια διεργασία, κάθε φυσική κίνηση του χρήστη ή κάθε γνωστικά ενιαία ενέργεια (που στα αγγλικά εκφράζεται με τον όρο chunk) του χρήστη;
- *Πώς αξιολογούμε τις όμοιες διεργασίες;* Πολλές φορές οι ίδιες διεργασίες εμφανίζονται σε διαφορετικά σημεία της χρήσης και πολλές φορές έχουν και διαφορετικό νόημα, ανάλογα με τις συγκεκριμένες συνθήκες. Θα θεωρήσουμε ότι αξιολογήθηκαν ή θα τις ξανααξιολογήσουμε υπό το νέο πρίσμα των παραμέτρων;

Η απάντηση που δίνουν όλοι οι συγγραφείς στη βιβλιογραφία και που ασπάζεται επίσης αυτή η διατριβή είναι ότι αυτές οι αποφάσεις παίρνονται πάντα σε συνάρτηση με το σύστημα υπό αξιολόγηση. Πρέπει να τονιστεί όμως σ’ αυτό το σημείο ακόμα μια φορά ότι οι παράμετροι αυτοί πρέπει να λαμβάνονται υπ’ όψιν κατά το σχεδιασμό της αξιολόγησης, άλλως διακινδυνεύουμε αναξιόπιστα αποτελέσματα.

γ. Ετοιμάζουμε τα ερωτηματολόγια ή το άλλο απαραίτητο υλικό καταγραφής της αξιολόγησης. Για παράδειγμα, αν ετοιμάζουμε συνεντεύξεις, πρέπει να αποφασίσουμε τη μορφή τους (δομημένες ή όχι) και να ετοιμάσουμε τις ερωτήσεις.

Το σημείο αυτό είναι μάλλον αυτονόητο, μιας που είναι αυτό που εννοούν οι περισσότεροι όταν αναφέρονται στην προετοιμασία της αξιολόγησης. Όπως όμως βλέπουμε, είναι μόνο ένα από τα στάδια της προετοιμασίας μιας αξιολόγησης.

δ. Ετοιμάζουμε έναν “Οδηγό για τους αξιολογητές” ή για τους χρήστες όπου περιγράφουμε πως πρέπει να αξιολογήσουν ή να συμπεριφερθούν. Αυτή η διατριβή προτείνει μάλιστα μια πιο ολοκληρωμένη προσέγγιση που θα εξηγηθεί στα επόμενα κεφάλαια: περιγράφονται με συντομία τόσο οι βασικές αρχές των αξιολογήσεων όσο και η συγκεκριμένη μεθοδολογία που ακολουθείται, παράλληλα με οδηγίες για το τι πρέπει να κάνουν οι αξιολογητές, έτσι ώστε αυτοί να αισθάνονται άνετα και να βοηθηθούν προς την κατεύθυνση που θέλουμε να οδηγήσουμε την αξιολόγηση. Ένα από τα συμπεράσματα αυτής της διατριβής είναι ότι το σημείο αυτό δεν θα μπορούσε να είχε παραλειφθεί στις συγκεκριμένες έρευνες που έγιναν, αντιθέτως, από τα σχόλια των αξιολογητών μας διαπιστώσαμε την ανάγκη ύπαρξής του φυλλαδίου αυτού.

ε. Επιλογή των ποιοτικών ή ποσοτικών δεδομένων που πρέπει να συλλεχθούν και να επεξεργαστούν. Μπορεί να τα χρειαζόμαστε όλα ή κάποια μόνο από αυτά.

Ποσοτικά:

1. *Μετρήσεις απόδοσης* (Performance measures)
2. *Μετρήσεις χρόνων* (Counting time)
3. *Επιδόσεις ποιότητας* (Quality scores)
 - *Συχνότητα χρήσης συγκεκριμένων εργαλείων* (Frequency of tools usage). Αυτή τη μέτρηση μπορεί κανείς να θέλει να την κάνει για δύο λόγους. Για να καταγράψει τη στρατηγική που χρησιμοποιεί ο κάθε χρήστης για την ολοκλήρωση της συγκεκριμένης διεργασίας, οπότε μπορεί να αποφασίσει κανείς για χρήση ή όχι εναλλακτικών (alternatives) και εξωτερικών (auxiliary) διεργασιών, ή για να πάρει κανείς ενδείξεις για τη χρησιμότητα (utility) του συγκεκριμένου εργαλείου κατά τη χρήση της διεργασίας.
 - *Κλίμακα βαθμολόγησης* τύπου Likert, όπου ο χρήστης-αξιολογητής θα αποτιμήσει αριθμητικά την υπό συζήτηση πράξη.

Ποιοτικά:

- ◆ *Κλειστές ερωτήσεις*, όπου ο χρήστης πρέπει να επιλέξει σε κάθε ερώτηση από ένα μικρό πλήθος δυνατών επιλογών την πλησιέστερη προς την άποψή του. Τα ερωτηματολόγια αυτού του τύπου μοιάζουν πολύ με ασκήσεις που πρέπει να εκτελέσει ο ερωτώμενος και ουσιαστικά ακολουθούν την ίδια φιλοσοφία. Για παράδειγμα οι τύποι (κλειστών) ασκήσεων που προτείνει ο Race (1993) στα πλαίσια της Ανοικτής Εκπαίδευσης θα μπορούσαν επίσης να είναι τύποι ερωτήσεων αξιολόγησης. Γενικά όμως μπορούμε να διακρίνουμε τους εξής τύπους κλειστών ερωτήσεων:
 - Ναι / όχι / δεν ξέρω
 - Πολλαπλών επιλογών (multiple choice) μονής επιλογής. Ο χρήστης επιλέγει μόνο μία.
 - Πολλαπλών επιλογών πολλαπλής επιλογής. Ο χρήστης μπορεί να επιλέξει περισσότερες από μία απαντήσεις.
 - Αντιστοίχιση. Πολλές ερωτήσεις με πολλές απαντήσεις. Ο χρήστης αντιστοιχεί κατά την κρίση του.
 - Σωστό / Λάθος
 - Κλίμακα τύπου Likert.
 - Σημασιολογικά διαφοροποιημένη κλίμακα, που ουσιαστικά αποτελεί βελτίωση της κλίμακας Likert.
 - Συμπλήρωση κενού. Λείπει η λέξη που χαρακτηρίζει την οντότητα.
 - Σωστή ακολουθία. Ο χρήστης βάζει στη σειρά χαρακτηρισμούς.
 - Εντοπισμός λάθους.

Στο *Παράρτημα 1*, δίνεται ένα πρότυπο ερωτηματολόγιο που περιέχει παραδείγματα αυτών των κατηγοριών ερωτήσεων.

- ◆ *Ανοικτές ερωτήσεις*, όπου ο χρήστης-αξιολογητής καλείται να διατυπώσει την άποψή του. Και οι ανοικτές ερωτήσεις υπακούουν σε κάποιους κανόνες. Για παράδειγμα, δεν ρωτάμε ποτέ «σας άρεσε η διασύνδεση, έτσι;» αλλά «τί σκέφτεστε για τη διασύνδεση;»

στ. Ετοιμασία του υλικού (hardware) και όποιου άλλου ειδικού εξοπλισμού είναι απαραίτητος για τη διεξαγωγή της αξιολόγησης. Το στάδιο αυτό μπορεί να κρύβει εκπλήξεις. Πολύ συχνά αφήνεται τελευταίο με την υπερβολική και αδικαιολόγητη εμπιστοσύνη ότι όλα θα δουλέψουν τέλεια. Η εμπειρία έχει δείξει ότι συνήθως κάτι δεν θα δουλέψει. Πρέπει λοιπόν να προγραμματιστεί και αυτό το στάδιο στη φάση της προετοιμασίας, ώστε να εξασφαλίσουμε την ομαλή διεξαγωγή της συνεδρίας. Μεγάλη βοήθεια εδώ θα μας δώσει η διεξαγωγή της πιλοτικής αξιολόγησης που εξηγείται παρακάτω.

ζ. Ετοιμασία βοηθητικού υλικού που μερικές φορές είναι απαραίτητο για τη διεξαγωγή της αξιολόγησης, πχ. χάρτες, διαγράμματα, οπτικοί σελιδοδείκτες (optical bookmarks) που θα χρησιμοποιήσουν οι αξιολογητές ή θα χρειαστούν κατά τη διεξαγωγή της συνεδρίας αξιολόγησης. Το στάδιο αυτό εξαρτάται ισχυρά από τη μεθοδολογία που θα ακολουθήσουμε. Εδώ δεν συμπεριλαμβάνεται η αναπαραγωγή ή φωτοτύπιση των ερωτηματολογίων που θα χρησιμοποιηθούν στη διάρκεια της συνεδρίας, καθώς θεωρείται ότι καλύφθηκε παραπάνω. Υπάρχει όμως σχεδόν πάντα βοηθητικό υλικό που την ώρα της συνεδρίας θα ήθελε κανείς να έχει στη διάθεσή του και πρέπει να προβλεφθεί σ' αυτό το στάδιο. Για παράδειγμα, αν η διάρκεια της συνεδρίας προβλέπεται μεγάλη, πρέπει να φροντίσουμε και για τα αναψυκτικά και τους καφέδες που θα προσφερθούν στο διάλειμμα.

4. Πιλοτική αξιολόγηση

Καλό είναι να γίνεται πάντα μια πιλοτική αξιολόγηση για να ρυθμίζονται οι τελευταίες λεπτομέρειες (πχ.στο ερωτηματολόγιο) και να γίνεται μια αποτίμηση του χρόνου που θα χρειαστεί. Η αξία της πιλοτικής αξιολόγησης έχει αναγνωριστεί από πολλούς ερευνητές (Shneiderman, 1998; Preece, 1994). Η πιλοτική αυτή αξιολόγηση θεωρείται από πολλούς ερευνητές σαν μέρος της κύριας αξιολόγησης. Οι Lewis & Rieman (1994) μάλιστα συνιστούν να γίνει δύο φορές: μία με τους συναδέλφους για να βρεθούν τα μεγάλα λάθη (bugs) και μία με κάποιους πραγματικούς χρήστες, όπου θα ελέγξουμε μικρότερα θέματα, όπως για παράδειγμα, το αν δουλεύει σωστά η πολιτική βοήθειας που ακολουθούμε ή αν οι οδηγίες μας προς τους αξιολογητές ήταν σαφείς. Ανακεφαλαιώνοντας, η πιλοτική αξιολόγηση μπορεί να:

- Μας πληροφορήσει για το χρόνο που θα χρειαστούμε
- Να αποκαλύψει λάθη στο υλικό που ετοιμάσαμε
- Να προπονήσει τους άγοντες την αξιολόγηση
- Να ανακαλύψει άλλα μη προβλεφθέντα θέματα

Γενικά, όπως είναι προφανές, η πιλοτική αξιολόγηση είναι μια *διαμορφωτική εμπειρική αξιολόγηση της αξιολόγησης* και σαν τέτοια δεν πρέπει να παραλείπεται.

5. Άγοντας την αξιολόγηση

Εδώ υπάρχουν πολλοί παράγοντες που διαφοροποιούνται, οι οποίοι προφανώς δεν μπορούν να αναφερθούν όλοι, αλλά υπάρχει και ο χρυσός κανόνας που λέει ότι όσο καλύτερα προετοιμαστεί μια αξιολόγηση, τόσο καλύτερα θα διεξαχθεί. Υπάρχουν όμως κάποια θέματα της διαδικασίας της διεξαγωγής που πρέπει να έχουν προαποφασιστεί. Δεν μπορούν να αναφερθούν όλα, λόγω της μεγάλης ποικιλομορφίας των μεθοδολογιών, όμως μπορούμε να αναφέρουμε ενδεικτικά κάποια που στην ουσία γεφυρώνουν το στάδιο της προετοιμασίας με αυτό της διεξαγωγής της συνεδρίας.

- Στις μεθόδους που εμπλέκουν χρήστες το περιβάλλον αξιολόγησης πρέπει να είναι όσο πιο κοντά στην πραγματικότητα γίνεται. Συνιστάται (Nielsen, 1993b; Lewis & Rieman, 1994) να αποφεύγεται το εργαστήριο σαν χώρος για τέτοιου είδους αξιολογήσεις και να χρησιμοποιείται ο πραγματικός χώρος εργασίας των χρηστών, όπου είναι αυτό δυνατό.
- Πάντα πρέπει να γίνεται σαφές ότι δεν αξιολογούνται οι αξιολογητές, αλλά το σύστημα. Γενικότερα πρέπει να λαμβάνονται υπ' όψιν και τα άλλα ηθικά θέματα της αξιολόγησης, όπως η προστασία των στοιχείων των αξιολογητών, η συμπεριφορά μας προς αυτούς, το δικαίωμά τους να αποχωρήσουν όποτε θελήσουν και άλλα τέτοια σχετικά θέματα που προκύπτουν από το γεγονός της εμπλοκής στη διαδικασία ανθρωπίνων όντων.
- Επιλέγουμε μέθοδο διεξαγωγής:
 - α) *Ατομική* (ένας αξιολογητής τη φορά). Αυτή η μέθοδος χρησιμοποιείται όταν η διαδικασία πρέπει να ακολουθηθεί κατά βήμα και με προσοχή και να καταγραφούν οι αντιδράσεις και απόψεις του κάθε αξιολογητή ξεχωριστά.
 - β) *Ομαδική* (όλοι οι αξιολογητές μαζί). Αυτή η μέθοδος χρησιμοποιείται αν μέσα στους στόχους μας είναι η σύγκριση των απόψεων των αξιολογητών με σκοπό την εξαγωγή κάποιων γενικών συμπερασμάτων.
- Αν η διάρκεια της αξιολόγησης προβλέπεται μεγάλη, την χωρίζουμε σε συνεδρίες (sessions), όχι μεγαλύτερες από 2 ώρες. Προβλέπουμε διάλειμμα και τους απαραίτητους καφέδες και αναψυκτικά για το διάλειμμα.

6. Επεξεργασία των αποτελεσμάτων

Σύμφωνα με τους Kantner & Rosenbaum (1997) τα αποτελέσματα των αξιολογήσεων οργατώνονται σε τέσσερις κατηγορίες:

- *Υποστήριξη των διεργασιών χρήστη (user task support)*. Έχει ήδη προαναφερθεί ότι ο χρηστοκεντρικός σχεδιασμός θεωρείται σήμερα σαν ο ακρογωνιαίος λίθος του usability engineering. Συνεπώς το να μπορεί ο χρήστης να εκτελέσει τις διεργασίες που θέλει μέσα στο σύστημα είναι πρωταρχικός στόχος.
- *Συμπεριφορά της διασύνδεσης χρήστη (user interface behaviour)*. Προς το σκοπό αυτό η διασύνδεση θα πρέπει να διευκολύνει το χρήστη με το να υπακούει στις αρχές της αισθηθαιναικότητας και της διαφάνειας, όπως ήδη είναι γνωστό.
- *Παρουσίαση (presentation)*. Για να πετύχει αυτούς τους στόχους, θα πρέπει η διασύνδεση επίσης να παρουσιάζει την πληροφορία με τον τρόπο που περιμένει ο χρήστης και όχι για παράδειγμα να τον ξαφνιάζει.
- *Ορολογία (terminology)*. Και τέλος μέσα στην έννοια του τρόπου που περιμένει ο χρήστης είναι και η ορολογία που ανταποκρίνεται στο συγκεκριμένο περιβάλλον και η οποία πρέπει να είναι πλήρως κατανοητή στο χρήστη.

Η επεξεργασία των ποιοτικών και ποσοτικών δεδομένων που προέκυψαν τελικά από την αξιολόγηση αποσκοπεί στη μετατροπή τους σε συμπεράσματα και συστάσεις για βελτιώσεις. Η ανάλυση των δεδομένων γενικά προϋποθέτει ακριβείς στατιστικές και μεθοδολογικές γνώσεις καθώς και θεωρητικές γνώσεις στο αντικείμενο αξιολόγησης.

Τα ποσοτικά δεδομένα επεξεργάζονται συνήθως με στατιστικές μεθόδους (μέσοι όροι, τυπικές αποκλίσεις κλπ.). Ευνόητο είναι ότι και εδώ πρέπει να έχουμε κάνει τη σωστή προετοιμασία. Η κωδικοποίηση για παράδειγμα πρέπει να γίνεται με κάποια αρχή και συνέπεια. Για παράδειγμα, όταν χρησιμοποιούνται κλίμακες διαβάθμισης, οι κωδικοί πρέπει να έχουν την ίδια φορά ή αντιστοιχία σε όλες τις ερωτήσεις ή μεταβλητές.

Αντίθετα τα ποιοτικά δεδομένα δίνουν ποιοτικά καλύτερα αποτελέσματα, όπως όμως προαναφέρθηκε, η επεξεργασία τους είναι πολύ πιο δύσκολη. Μπορούμε να προσπαθήσουμε να ποσοτικοποιήσουμε κάποιες από αυτές τις ερωτήσεις. Στις ανοικτού τύπου ερωτήσεις, τα δεδομένα μπορούν να ποσοτικοποιηθούν για παράδειγμα με μια διαδικασία κωδικοποίησης. Όμως υπάρχουν και μέθοδοι που επιτρέπουν να αναλυθεί το περιεχόμενό τους ποιοτικά. (Μακράκης, 1999)

Πρέπει να αποφεύγουμε τόσο τις υπεραπλουστεύσεις όσο και τις υπερδιογκώσεις των αποτελεσμάτων της ανάλυσης. Πρέπει επίσης να συζητηθούν τα ερωτήματα που έμειναν αναπάντητα ή για διάφορους λόγους δεν ελήφθησαν υπόψη στην έρευνα-αξιολόγηση και τα οποία χρήζουν περαιτέρω διερεύνησης. Πρέπει ακόμα να γίνεται προσπάθεια διείσδυσης σε βάθος στα αποτελέσματα της έρευνας-αξιολόγησης, έτσι ώστε να συλλαμβάνονται οι ακριβείς ερμηνείες και οι συνέπειές τους σε θεωρητικό και πρακτικό επίπεδο.

Στο σημείο αυτό ολοκληρώνουμε τη μελέτη μας σχετικά με τα στάδια της αξιολόγησης. Στη συνέχεια δίνονται δύο πίνακες που συνοψίζουν τα παρουσιασθέντα μέχρι τώρα.

Ο πίνακας αυτός παρουσιάζει με εποπτική μορφή μια σύνοψη των διαφόρων φάσεων:

Μέθοδος	Επιλογή Μεθόδου	Προετοιμασία	Εκτέλεση	Αποτίμηση
Analytic Evaluation (αναλυτική αξιολόγηση)	Αν είναι νωρίς στον κύκλο σχεδιασμού (design cycle), δηλαδή στον σχεδιασμό ή στον καθορισμό των προδιαγραφών συστήματος (system specifications)	<ul style="list-style-type: none"> • Ετοιμάζουμε το πρωτότυπο προς αξιολόγηση • Ετοιμάζουμε το υλικό καταμέτρησης 	Ανάλυση κάθε διεργασίας και υπολογισμοί χρόνων Αξιολόγηση και εναλλακτικών διαδρομών.	Στατιστική αποτίμηση. Δίνει αποτελέσματα με νούμερα σχετικά με την επίδοση του χρήστη.
Expert evaluation (αξιολόγηση των ειδικών)	Μόλις ετοιμαστεί ένα πρωτότυπο	<ul style="list-style-type: none"> • Επιλέγουμε αξιολογητές • Καθορίζουμε τις διεργασίες • Ετοιμάζουμε τα ερωτηματολόγια • Ετοιμάζουμε τον οδηγό των αξιολογ/τών 	Συνήθως χρησιμοποιείται η δομή Στόχος (goal) → διεργασία (task) → ενέργεια (action)	<ul style="list-style-type: none"> • Είναι ισοδύναμες οι απαντήσεις όλων των αξιολογητών; • Είναι ισοβαρείς όλες οι ερωτήσεις; • Καθορισμός των “γενικών” ερωτήσεων που θα απαντηθούν.
Observational Empirical evaluation: (παρατηρητική αξιολόγηση)	<ul style="list-style-type: none"> • Αν έχουμε έτοιμη beta version • Αν στοχεύουμε σε ομάδα ειδικών χρηστών (special interest user group) 	<ul style="list-style-type: none"> • Επιλέγουμε αξιολογητές • Καθορίζουμε τις διεργασίες • Καθορίζουμε τις παραμέτρους που θα καταγράψουν την επίδοση 	Μετά από μία μικρή περίοδο γνωριμίας με το σύστημα ζητείται από κάθε αξιολογητή να ολοκληρώσει τις διεργασίες. Δίνεται σαν ανώτερος χρόνος για την κάθε μια κάποιος χρόνος που εκτιμήθηκε σαν αρκετός.	<ul style="list-style-type: none"> • Καθορισμός των παραμέτρων αποτίμησης • Στατιστική επεξεργασία των ποσοτικών δεδομένων
Survey Empirical evaluation: (καταμετρητική αξιολόγηση)	Ομοίως	<ul style="list-style-type: none"> • Επιλέγουμε αξιολογητές • Καθορίζουμε τις διεργασίες • Ετοιμάζουμε τα ερωτηματολόγια 	Συμπλήρωση των ερωτηματολογίων ή διεξαγωγή της συνέντευξης, όπως προετοιμάστηκε.	Στατιστική ή όχι επεξεργασία. Οι υποκειμενικές γνώμες μπορούν να ομαδοποιηθούν.
Experimental evaluation (πειραματική αξιολόγηση)	<ul style="list-style-type: none"> • Αν θέλουμε να αξιολογήσουμε διασύνδεση με πολλές παραμέτρους που έχει πολλαπλές λειτουργίες και μορφές • Αν στοχεύουμε σε ειδική ομάδα χρηστών 	<ul style="list-style-type: none"> • Καθορισμός των παραμέτρων που θα χειριστούμε • Ετοιμασία του απαραίτητου εξοπλισμού 	Χρήστες εργάζονται επαναληπτικά με το σύστημα κάτω από διαφορετικές συνθήκες εργασίας.	Στατιστικά συγκρίσιμα αποτελέσματα.

Δίνεται τέλος και ο επόμενος πίνακας, που συνοψίζει τις μεθοδολογίες σύμφωνα με μια πρώτη προσέγγιση ως προς τα υπέρ και τα κατά που αναφέρθηκαν έως τώρα:

Μέθοδος	Περιγραφή	Υπέρ/Κατά	Πότε χρησιμοποιείται
Analytic Evaluation (αναλυτική αξιολόγηση)	Χρησιμοποιεί μια επίσημη ή ημιεπίσημη περιγραφή της διασύνδεσης για να προβλέψει την επίδοση του χρήστη.	Δεν χρειάζεται πρωτότυπο ή χρήστες. Όμως έχει περιορισμένη εστίαση και δεν δίνει λύσεις για επανασχεδιασμό.	Μπορεί να γίνει νωρίς στο design cycle (απασχολεί λίγα resources)
Expert evaluation (αξιολόγηση των ειδικών)	Ειδικοί κρίνουν το σύστημα και προσπαθούν να βρουν προβλήματα στη χρήση παίζοντας το ρόλο λιγότερο έμπειρων χρηστών.	Φτηνή και αποδοτική μέθοδος, μιας και λίγοι ειδικοί μπορούν να ανακαλύψουν σημαντικά προβλήματα.	Η μέθοδος μπορεί να εφαρμοστεί σε όλο το στάδιο σχεδιασμού και κατασκευής.
Empirical evaluation: a) Observational (παρατηρητική αξιολόγηση)	Συλλέγουμε δεδομένα από τη συμπεριφορά των χρηστών κατά την χρήση του συστήματος.	Η ανάλυση στοιχείων που ακολουθεί τη συλλογή δεδομένων είναι χρονοβόρα και καταναλώνει πόρους, όμως δίνει ποιοτικά ενδιαφέροντα αποτελέσματα. Η μέθοδος μπορεί να αποβεί πολύ «φασαριατζίδικη» και να επηρεάσει την δραστηριότητα και την αποδοτικότητα των χρηστών.	Χρησιμοποιείται συνήθως σε πρωτότυπα που έχουν φτάσει σε σημαντικό βαθμό ολοκλήρωσης και είναι σχεδόν έτοιμα προς χρήση. Συνεπώς ακριβή μέθοδος σε σύγκριση με άλλες.
Empirical evaluation: b) Survey (καταμετρητική αξιολόγηση)	Χρησιμοποιούνται συνεντεύξεις ή ερωτηματολόγια για να εξαχθούν οι υποκειμενικές γνώμες των χρηστών για το interface.	Οι συνεντεύξεις χρησιμοποιούνται με ένα περιορισμένο γκρουπ χρηστών και μπορεί να είναι χρονοβόρες. Τα ερωτηματολόγια μπορούν να χρησιμοποιηθούν με μεγάλα γκρουπ χρηστών και είναι εύχρηστα.	Ομοίως
Experimental evaluation (πειραματική αξιολόγηση)	Ο αξιολογητής μπορεί να χειριστεί κάποιες παραμέτρους που συνδέονται με το σχεδιασμό του interface και να τις μεταβάλλει. Μελετάει την επίδραση στην απόδοση των χρηστών κατά τις μεταβολές και σε διάφορα θέματα.	Προϋποθέτει καλή γνώση πειραματικής μεθοδολογίας, είναι χρονοβόρο και χρειάζεται πόρους. Έχει περιορισμένο εύρος αξιολόγησης, αλλά δίνει καλά αποτελέσματα σε καλά εστιασμένα ερωτήματα.	Χρησιμοποιείται συνήθως σε πλήρως εξελιγμένα προϊόντα.

1.8. Μια αναλυτικότερη προσπάθεια ταξινόμησης

Στη συνέχεια γίνεται μια προσπάθεια να ταξινομηθούν πιο αναλυτικά οι περισσότερες από τις μεθοδολογίες αξιολόγησης που συναντώνται στη βιβλιογραφία σήμερα. Η ταξινόμηση αυτή συναντάται σε πολλούς συγγραφείς με περισσότερη ή λιγότερη επιτυχία, για παράδειγμα ο Nielsen (1993b, σελ 224) δίνει ένα τέτοιο περιληπτικό πίνακα, ενώ υπάρχουν ειδικά βιβλία πάνω στο θέμα, όπως για παράδειγμα το Nielsen & Mack (1994). Η προσπάθεια που γίνεται συνεπώς εδώ έχει πιο περιορισμένο σκοπό, αφ' ενός τη συγκέντρωση των απόψεων των σημαντικότερων συγγραφέων του χώρου και αφ' ετέρου την παράθεση των συγκριτικών πλεονεκτημάτων και μειονεκτημάτων της κάθε προσέγγισης, καθώς και τις περιπτώσεις στις οποίες συνιστάται η εφαρμογή κάθε προσέγγισης. Δεν θα δώσουμε αναλυτικές περιγραφές αυτών των μεθοδολογιών, παρά μόνο σύντομες, γιατί κάτι τέτοιο θα επεξέτεινε πολύ το μέγεθος αυτού του κεφαλαίου. Ο ενδιαφερόμενος αναγνώστης μπορεί να βρει λεπτομέρειες για κάθε μία από αυτές τις μεθόδους στη συνιστώμενη κατά περίπτωση βιβλιογραφία που συνοδεύει τη συγκεκριμένη μέθοδο.

Η ταξινόμηση αυτή κατηγοριοποιεί τις μεθοδολογίες με βάση το «διακρίνον (identifying) χαρακτηριστικό», εκείνο δηλαδή το χαρακτηριστικό της αξιολόγησης που καθορίζει έντονα τη φυσιογνωμία της. Ο αναγνώστης θα παρατηρήσει ασφαλώς ότι κάποιες από τις αναφερθείσες μεθοδολογίες θα μπορούσαν ενδεχομένως να κατηγοριοποιηθούν και αλλιώς. Συνεπώς η έρευνα σ' αυτό το σημείο δεν θεωρείται ολοκληρωμένη. Επίσης προφανώς μπορούμε να έχουμε μίξη από κάποιες από αυτές τις μεθοδολογίες όπως και συνδυασμούς αυτών. Για παράδειγμα μπορούμε να συνδυάσουμε σχεδόν όλες τις εμπειρικές μεθόδους με επακόλουθες συνεντεύξεις των χρηστών. Επίσης αυτή η διατριβή παρουσιάζει στο «*Κεφάλαιο 3: Η Συνδυαστική Αξιολόγηση*» μια προσέγγιση συνδυασμού εμπειρικής μεθόδου με μέθοδο βασισμένη σε ειδικούς.

1. Αναλυτικές μέθοδοι (analytic evaluation)

Είναι διαμορφωτικές – ποσοτικές

Χρησιμοποιούν μια «επίσημη» ή ημι-επίσημη (semi-formal) περιγραφή της διασύνδεσης για να προβλέψουν την επίδοση του χρήστη σε φυσικές και αντιληπτικές ενέργειες που πρέπει να γίνουν. Δεν χρειάζονται ακριβή πρωτότυπα και εμπλοκή χρηστών. Όμως έχουν περιορισμένη εστίαση (narrow focus) και δεν δίνουν λύσεις για επανασχεδιασμό, αφού ουσιαστικά βασίζονται σε υποθέσεις σχετικά με τις αντιδράσεις του χρήστη ή σε μετρήσεις χρονικών επιδόσεων. Γίνονται συνήθως νωρίς στον κύκλο σχεδιασμού (design cycle), αλλά μπορούν να εφαρμοστούν και αργότερα σε συγκεκριμένα σημεία της διασύνδεσης με σκοπό τη βελτίωσή τους.

☞ **Ανάλυση Επιπέδου Πληκτρολόγησης (Keystroke-Level Analysis).**

Βασίζεται στον υπολογισμό χρόνων κατά την πληκτρολόγηση, ή τη διαχείριση του ποντικιού. Η μέθοδος πρωτοπαρουσιάστηκε από τους Card et al. (1983) σαν «Μοντέλο Επιπέδου Πληκτρολόγησης – Keystroke Level Model» και βασίστηκε στις θεωρίες γνωστικής ψυχολογίας του Alan Newel (που είναι ένας από τους συγγραφείς) και του Herbert Simon. Η προσέγγιση των Card et al. (1983) είναι γνωστή σαν GOMS – Goals, Operators, Methods and Selection. Η μέθοδος περιγράφεται αναλυτικά στον Αβούρη (2000, p.172) και στους Lewis & Rieman (1994, p.37).

Υπέρ	Κατά
<ul style="list-style-type: none">• Ακριβής πρόβλεψη επιδόσεων του χρήστη• Εφαρμόζεται νωρίς στο σχεδιαστικό κύκλο• Αποδίδει καλά σε περίπλοκα interfaces• «Σκληρά» νούμερα, στατιστικά επεξεργάσιμα	<ul style="list-style-type: none">• Δύσκολη μέθοδος, απαιτεί προσπάθεια• Δεν εφαρμόζεται πάντα• Δεν αξιολογεί όλες τις πλευρές του interface, αλλά μόνο την επίδοση• Προβλέπει μόνο τις επιδόσεις του έμπειρου χρήστη• Απαραίτητη μια σειρά διεργασιών & ενεργειών (διεργασιοκεντρική)

Συνιστάται: 1) σε ειδικές περιπτώσεις, όταν το μεγάλο κόστος της σε προσπάθεια αποπληρώνεται με μεγάλο κέρδος 2) σε τμήματα της διασύνδεσης που οι χρήστες θα προσβαίνουν συχνά

☞ **«Πρόχειρη» Ανάλυση («Back-of-the-Envelope»).** (Lewis & Rieman, 1994, p.41-42).

Αποφεύγει τις πολύ αναλυτικές περιγραφές και στοχεύει να ρίξει μια γρήγορη ματιά στην ολοκληρωμένη εικόνα. Καθορίζονται οι ενέργειες και γίνονται πρόχειροι υπολογισμοί λαμβάνοντας υπ' όψιν: πόσες ενέργειες χρειάζονται, 2-3 δευτερόλεπτα για κάθε ενέργεια, αν υπάρχουν συντομότεροι δρόμοι, τί πρέπει να μάθει ο χρήστης κλπ.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Διατηρεί τη γενική εικόνα. • Απασχολεί λιγότερους πόρους • Γίνεται νωρίς στο σχεδιαστικό κύκλο • Αποδίδει καλά σε περίπλοκα interfaces 	<ul style="list-style-type: none"> • Δεν είναι ακριβής. • Δεν εφαρμόζεται πάντα • Δεν αξιολογεί όλες τις πλευρές του interface, αλλά μόνο την επίδοση. • Απαραίτητη μια σειρά διεργασιών & ενεργειών (διεργασιοκεντρική)

Συνιστάται για να αποφασίσουμε αν θα προσθέσουμε ή όχι δυνατότητες στη διασύνδεση, ή και σε όλο το σύστημα.

☞ **Συνεργατικός Έλεγχος (Synergy Review).** (Nielsen, 1993b, σελ. 236).

Περισσότερα από ένα προϊόντα αξιολογούνται συγχρόνως εσωτερικά, στον κατασκευαστή, από ένα μίγμα αξιολογητών από διαφορετικούς τομείς, σε ένα από τα πρώτα στάδια σχεδιασμού του. Οι αξιολογητές προσπαθούν να βρουν κοινά και επικαλυπτόμενα σημεία ανάμεσα στα προϊόντα, τα οποία εν συνεχεία θα σχεδιαστούν μια φορά για όλα τα προϊόντα.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Συγκρίνει τις δυνατότητες περισσότερων προϊόντων • Μειώνει το κόστος • Προσθέτει συνέπεια στις διασυνδέσεις • Ευνοεί την ανταλλαγή ιδεών μεταξύ των σχεδιαστών 	<ul style="list-style-type: none"> • Όχι γενικά εφαρμόσιμη, αφορά μόνο έναν κατασκευαστή κάθε φορά. • Πολλές προτάσεις ίσως να μην είναι υλοποιήσιμες • Απαιτεί περισσότερα προϊόντα σε αρχικό στάδιο σχεδιασμού.

Συνιστάται όταν υπάρχουν περισσότερα του ενός προϊόντα προς κατασκευή

☞ **Έλεγχος Χαρακτηριστικών (Feature Inspection).** (Nielsen & Mack, 1994; Nielsen, 2001a).

Η μέθοδος αυτή παρουσιάζει την αλληλουχία των χαρακτηριστικών που χρησιμοποιούνται για να ολοκληρωθούν οι συνηθισμένες διεργασίες των χρηστών και προσπαθεί να αξιολογήσει το προτεινόμενο σετ χαρακτηριστικών, όπως για παράδειγμα υπερβολικά μεγάλες αλληλουχίες ενεργειών, περίπλοκα "μπελαλίδικα-cumbersome" βήματα, βήματα που δεν είναι φυσικά στην εκτέλεσή τους και βήματα που απαιτούν εκτεταμένη γνώση / εμπειρία για να ολοκληρωθεί το σετ των βημάτων αυτών.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Δίνει λεπτομερείς αναφορές για τις αξιολογηθείσες διεργασίες • Βρίσκει λεπτομέρειες που ξεφεύγουν από το σχεδιασμό 	<ul style="list-style-type: none"> • Δεν αξιολογεί όλη τη διασύνδεση • Χάνεται η εποπτεία του όλου

Συνιστάται όταν θέλουμε να βελτιώσουμε συγκεκριμένα σημεία στη διασύνδεση

☞ **Επιθεώρηση Προδιαγραφών (Standards Inspection).** (Nielsen & Mack, 1994; Nielsen, 2001a)

Το σύστημα συγκρίνεται με λίστες που αναφέρουν τόσο συμβάσεις (standards) του Διεθνούς Οργανισμού Τυποποίησης (ISO – International Standards Organization) (Brooke et al., 1990), όσο και της Ευρωπαϊκής Ένωσης (Stewart, 1990), όπως επίσης και των μεγάλων εταιρειών κατασκευής λογισμικού (Berry, 1988; Good, 1989; Tognazzini, 1989; Wolf, 1989). Τους ελέγχους αυτούς κάνουν συνήθως ειδικοί στους τομείς αυτούς. Η βασικότερη έννοια εδώ είναι η *συνέπεια*.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Εξασφαλίζει ευχρηστία για το χρήστη και χαμηλή καμπύλη εκμάθησης • Εξασφαλίζει μακροπρόθεσμα μείωση κόστους για την εταιρεία • Μπορεί να γίνει από εσωτερικούς αξιολογητές 	<ul style="list-style-type: none"> • Η συμμόρφωση με τις συμβάσεις μπορεί να σημαίνει βραχυπρόθεσμα έξοδα για την εταιρεία • Η συμμόρφωση με μια σύμβαση μπορεί να εμποδίζει την εξέλιξη και να αποθαρρύνει τους σχεδιαστές από το να δημιουργούν

Συνιστάται όταν έχουμε εμπορικό προϊόν ευρείας αποδοχής.

☛ **Επιθεώρηση Συνέπειας (Consistency Inspection).** (Wixon et al., 1994; Nielsen & Mack, 1994; Nielsen, 2001a).

Το interface αξιολογείται με ανα χείρας μια λίστα in-house προδιαγραφών, προδιαγραφών δηλαδή που ακολουθούν όλα τα προϊόντα του συγκεκριμένου κατασκευαστή, και ελέγχεται ως προς τη συμμόρφωσή του με αυτές τις οδηγίες. Συνήθως ελέγχονται περισσότερα από ένα προϊόντα συγχρόνως και οι παρόμοιες λειτουργίες συγκρίνονται μεταξύ τους από τους σχεδιαστές όλων αυτών των προϊόντων για να διαπιστωθεί αν όλα δουλεύουν με τον ίδιο τρόπο, όσο αυτό είναι φυσικά εφικτό.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Δίνει ενιαία όψη και αίσθηση (look and feel) και ελαχιστοποιεί την καμπύλη εκμάθησης • Μπορεί να γίνει από εσωτερικούς αξιολογητές 	<ul style="list-style-type: none"> • Όχι γενικά εφαρμόσιμη, αφορά μόνο έναν κατασκευαστή κάθε φορά. • Πολλές ασυνέπειες ίσως να μην αίρονται, λόγω της ειδικής φύσης κάθε προϊόντος

Συνιστάται όταν ζητάμε εξοικονόμηση πόρων και επιδίωξη συνέπειας σε σειρά προϊόντων

2. Μέθοδοι βασισμένες σε ειδικούς (expert evaluation)

Είναι διαμορφωτικές – ποσοτικές/ποιοτικές

Ειδικοί, συνήθως του χώρου του HCI ή/και ειδικοί στο συγκεκριμένο πεδίο της εφαρμογής, αλλά στην ιδανική περίπτωση και στους δύο τομείς, κρίνουν το σύστημα και προσπαθούν να βρουν προβλήματα στη χρήση του παίρνοντας το ρόλο λιγότερο έμπειρων χρηστών. Χρησιμοποιούνται συνήθως ερωματολόγια για να αξιολογηθεί μια σειρά ενεργειών και διεργασιών. Φτηνές και αποδοτικές μέθοδοι, σε σύγκριση με τις μεθόδους που εμπλέκουν χρήστες, μιας και λίγοι ειδικοί μπορούν να ανακαλύψουν σημαντικά προβλήματα. Όμως πρέπει να επιλεγούν προσεκτικά για να μην έχουν προκαταλήψεις (biases), αλλά και για να μπορέσουμε να πλησιάσουμε την πραγματική συμπεριφορά των χρηστών. Οι μέθοδοι μπορούν να εφαρμοστούν σε όλο σχεδόν τον κύκλο του σχεδιασμού, σε αρχικά προτότυπα (first prototypes) ή ακόμα και κατά τη διάρκεια του καθορισμού των χαρακτηριστικών του συστήματος (system specifications).

Οι μέθοδοι αυτές συναντώνται στη βιβλιογραφία και με τον όρο «Μέθοδοι Επιθεώρησης Ευχρηστίας – Usability Inspection Methods», αφού οι ειδικοί στην ουσία «επιθεωρούν» τη διασύνδεση.

☛ **Γνωστική Περιδιάβαση (Cognitive Walkthrough).**

Η μέθοδος προτάθηκε αρχικά από τους Lewis et al., (1990), ενώ περιγράφεται αναλυτικά στα Polson et al (1992) και Wharton et al. (1994). Στοχεύει στο να αξιολογήσει διασυνδέσεις περιστασιακής χρήσης (walk-up-and-use), όπου η επιτυχής χρήση με την πρώτη προσπάθεια είναι απαίτηση. Ειδικοί ακολουθούν μια σειρά προκαθορισμένων διεργασιών και ενεργειών στο interface και προσπαθούν να ανακαλύψουν προβλήματα που θα αντιμετωπίσουν οι υποψήφιοι χρήστες.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Φτηνή και εύκολη μέθοδος. • Αποδοτική 	<ul style="list-style-type: none"> • Απαραίτητοι αξιολογητές με εμπειρία. • Απαραίτητη μια σειρά διεργασιών & ενεργειών (διεργασιοκεντρική) • Αργή και βαρετή μέθοδος • Εφαρμόζεται μόνο σε ειδικά interfaces • Δεν αποκαλύπτει όλα τα προβλήματα

Συνιστάται όταν έχουμε διασυνδέσεις απευθυνόμενες σε περιστασιακούς (casual) χρήστες.

☞ **Γνωστική Διάσχιση (Cognitive Jogthrough).**

Η μέθοδος αποτελεί βελτίωση του walkthrough. Προτάθηκε από τους Rowley & Rhoades (1992), οι οποίοι πρόσθεσαν μια βιντεοκάμερα με σκοπό να επιταχύνουν την συνεδρία.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Φτηνή και εύκολη μέθοδος. • Αποδοτική • Γρηγορότερη διεξαγωγή. 	<ul style="list-style-type: none"> • Απαραίτητοι αξιολογητές με εμπειρία. • Απαραίτητη μια σειρά διεργασιών & ενεργειών (διεργασιοκεντρική) • Εφαρμόζεται μόνο σε ειδικά interfaces • Δεν αποκαλύπτει όλα τα προβλήματα

Συνιστάται όταν έχουμε διαθέσιμους ειδικούς του HCI και της γνωστικής επιστήμης

☞ **Γραφική Γνωστική Περιδιάβαση & Διάσχιση (Cognitive Graphical Walk- and Jogthrough)**

Η μέθοδος προτάθηκε από εμάς στο Karoulis et al. (2000) και αποτελεί βελτίωση των υπαρχουσών μεθοδολογιών με την προσθήκη διαγραμμάτων, όπου οι αξιολογητές μπορούν να καταγράψουν την αύξηση της εμπειρίας του χρήστη με το πέρασμα του χρόνου. Γίνεται έτσι η μέθοδος κατάλληλη και για την αξιολόγηση εκπαιδευτικών αλλά και γενικότερης χρήσεως περιβαλλόντων.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Φτηνή και εύκολη μέθοδος. • Αποδοτική • Κατάλληλη για πολλά είδη διασυνδέσεων • Αυξημένη αξιοπιστία • Υπόκειται σε στατιστική επεξεργασία 	<ul style="list-style-type: none"> • Η εμπειρία των αξιολογητών, ειδικά στον υπό αξιολόγηση χώρο είναι απαραίτητη • Απαραίτητη μια σειρά διεργασιών & ενεργειών (διεργασιοκεντρική) • Δεν αποκαλύπτει όλα τα προβλήματα

Συνιστάται για εκπαιδευτικές διασυνδέσεις και σαν γενικό εργαλείο αξιολόγησης

☞ **Πολυμορφική Περιδιάβαση (Pluralistic Walkthrough) (Bias, 1991)**

Η αξιολόγηση γίνεται από ένα γκρουπ αξιολογητών αποτελούμενο από αντιπροσωπευτικούς χρήστες, σχεδιαστές και ειδικούς του HCI. Αρχικά κάθε αξιολογητής αξιολογεί μόνος του βάσει κριτηρίων, ενώ στη συνέχεια συζητούνται ομαδικά οι διεργασίες.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Ο καθένας φέρνει τη δική του εμπειρία • Συνδυασμός ευρετικής και διεργασιοκεντρικής αξιολόγησης • Η συμβίωση σχεδιαστών και χρηστών επιτρέπει την πρόωρη συλλογή δεδομένων χρήσης σε αρχικά στάδια σχεδιασμού 	<ul style="list-style-type: none"> • Οι ειδικοί των υπολογιστών τείνουν να «κυριαρχούν» (dominate) • Δύσκολη και περίπλοκη προετοιμασία • Χαλαρή δομή αξιολόγησης

Συνιστάται όταν δεν έχουμε αρκετούς ειδικούς ή δεν έχουμε «διπλοειδικούς» ή μας λείπει εντελώς η εμπειρία από τον τομέα που απευθύνεται το σύστημα.

☞ **Ευρετική Αξιολόγηση (Heuristic Evaluation).**

Η ονομασία αυτής της μεθόδου έχει, όπως είναι προφανές, ελληνική ρίζα. Δεν είναι όμως γνωστό το γιατί οι εμπνευστές της μεθόδου Nielsen & Molich (1990) την ονόμασαν έτσι. Σύμφωνα με τους Lewis & Rieman (1994) η λέξη «heuristic» είναι συνώνυμη του «guideline – οδηγία, κριτήριο». Στην παρούσα διατριβή χρησιμοποιείται είτε ο αγγλικός όρος, είτε η απόδοσή του στα ελληνικά σαν «κριτήριο», αφού μια μετάφραση με βάση την έννοια «εύρημα» μάλλον σύγκριση θα προκαλούσε στους ελληνόφωνους.

Στη μεθοδολογία αυτή ειδικοί διατρέχουν τη διασύνδεση με τα κριτήρια ανά χείρας, προσπαθώντας να διαπιστώσουν προβλήματα που θα έχουν οι χρήστες στην πραγματική εργασία τους με το σύστημα. Ο μετέπειτα συνδυασμός των αναφορών των αξιολογητών σε μια μοναδική λίστα δίνει ένα γκρουπ προβλημάτων που είναι αντιπροσωπευτικό για το συγκεκριμένο interface.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Χρειάζεται λίγους πόρους. • Μπορεί να εφαρμοστεί σε οποιοδήποτε στάδιο του σχεδιασμού. • Μπορεί να ανακαλύψει ένα ευρύ φάσμα προβλημάτων • Βρίσκει όλα τα σοβαρά προβλήματα 	<ul style="list-style-type: none"> • Μεγάλη λίστα κριτηρίων δύσκολο να εφαρμοστεί. Μικρή λίστα, χάνει προβλήματα. • Διαφεύγουν τα προβλήματα που δεν αναφέρονται στη λίστα κριτηρίων • Απαιτεί μεγάλη εμπειρία των αξιολογητών στο HCI. • Είναι δύσκολη η εφαρμογή των γενικών κριτηρίων σε ειδικά interfaces.

Συνιστάται όταν έχουμε ειδικούς του HCI και σαν γενικό εργαλείο αξιολόγησης

☞ **Ευρετική Εκτίμηση (Heuristic Estimation).** (Nielsen & Mack, 1994; Nielsen, 2001a)

Αυτή η μέθοδος είναι μια παραλλαγή της Ευρετικής Αξιολόγησης στην οποία οι ειδικοί - αξιολογητές εκτιμούν την *σχετική* ευχρηστία δύο ή περισσότερων πρωτοτύπων με ποσοτικούς τρόπους, όπως την προβλεπόμενη απόδοση (performance) του χρήστη.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Διατηρεί την εικόνα του όλου • Καλή σχέση κόστους/απόδοσης • Μπορεί να συγκρίνει περισσότερα interfaces • Στατιστικά επεξεργάσιμα αποτελέσματα 	<ul style="list-style-type: none"> • Δεν βρίσκει με λεπτομέρεια προβλήματα • Δεν δίνει υποδείξεις για το κάθε interface ξεχωριστά • Οι αξιολογητές πρέπει να είναι ειδικοί σε όλα τα αξιολογούμενα interfaces

Συνιστάται όταν έχουμε δύο ή περισσότερα πρωτότυπα και θέλουμε να επιλέξουμε το καλύτερο ή να καταλήξουμε σε συνδυασμούς αυτών.

☞ **Συμμετοχική Ευρετική Αξιολόγηση (Participatory Heuristic Evaluation).** (Muller et al., 1998)

Αποτελεί επέκταση της κλασικής ευρετικής μεθοδολογίας. Προσθέτει έμπειρους χρήστες του τομέα που αναφέρεται το λογισμικό στην ομάδα των αξιολογητών σαν «ειδικούς τομέα – work-domain experts» και προσθέτει επίσης άλλα πέντε κριτήρια στην υπάρχουσα λίστα με στόχο να γίνει η όλη διαδικασία πιο *διεργασιοκεντρική*, αφού, όπως ισχυρίζονται οι συγγραφείς, η αρχική λίστα είναι πιο πολύ προσανατολισμένη προς το προϊόν (product-oriented), παρά στις διεργασίες των χρηστών.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Περισσότερο πλουραλιστική ομάδα αξιολογητών • Καλύπτει γνωστικά και τον υπό συζήτηση τομέα • Η λίστα κριτηρίων καλύπτει και άλλα θέματα (πχ. δεξιότητες χρήστη, διάδραση, ιδιωτικότητα, αλληλουχία ενεργειών) 	<ul style="list-style-type: none"> • Οι απόψεις των χρηστών περνούν μέσα από δομημένη διαδικασία • Οι χρήστες-αξιολογητές δυσκολεύονται σε προβλήματα που απαιτούν γνώση HCI • Περιορίζει το εύρος του σταδίου σχεδιασμού που μπορεί να εφαρμοστεί

Συνιστάται όταν διαθέτουμε χρήστες, γνώστες του τομέα που αναφέρεται το λογισμικό

☞ **Επίσημη Επιθεώρηση Ευχρηστίας (Formal Usability Inspection).** (Nielsen & Mack, 1994; Nielsen, 2001a)

Η προσέγγιση αυτή συνδυάζει ατομικές και ομαδικές αξιολογήσεις σε μια διαδικασία έξι βημάτων. Υπάρχουν σαφώς καθορισμένοι ρόλοι για τους αξιολογητές, ενώ η μέθοδος συνδυάζει στοιχεία και από την ευρετική αξιολόγηση και από την απλοποιημένη μορφή των γνωστικών περιδιαβάσεων (cognitive walkthroughs).

Υπέρ	Κατά
<ul style="list-style-type: none"> • Επεξεργάσιμα αποτελέσματα • Συνδυάζει πλεονεκτήματα από άλλες μεθόδους 	<ul style="list-style-type: none"> • Περίπλοκη διαδικασία • Αρκετοί πόροι • Έμπειροι αξιολογητές

Συνιστάται όταν θέλουμε μια "όλα σε ένα" αξιολόγηση.

3. Εμπειρικές μέθοδοι αξιολόγησης

Παρατηρητικές Εμπειρικές (observational empirical evaluation)

Είναι διαμορφωτικές/αποτιμητικές – ποιοτικές/ποσοτικές.

Συλλέγουμε δεδομένα από τη συμπεριφορά των χρηστών κατά την χρήση του συστήματος. Μπορούν να χρησιμοποιηθούν πολλές τεχνικές: άμεση παρατήρηση, βιντεοεγγραφή (video recording), καταγραφή με λογισμικό (software logging – η διάδραση του χρήστη με το σύστημα καταγράφεται αυτόματα) και “ομιλούντα πρωτόκολα – verbal protocols” (ζητείται από το χρήστη να σκέφτεται δυνατά και να εκφράζει τις παρατηρήσεις του και τις σκέψεις του). Η ανάλυση στοιχείων που ακολουθεί τη συλλογή δεδομένων είναι χρονοβόρα και καταναλώνει πόρους, όμως δίνει ποιοτικά ενδιαφέροντα αποτελέσματα. Οι μέθοδοι αυτές μπορούν να αποβούν πολύ «οχληρές» (φασαριατζίδικες) και να επηρεάσουν την δραστηριότητα και την αποδοτικότητα των χρηστών. Οι μέθοδοι χρησιμοποιούνται συνήθως σε πρωτότυπα που έχουν φτάσει σε σημαντικό βαθμό ολοκλήρωσης και είναι σχεδόν έτοιμα προς χρήση. Συνεπώς ακριβές μέθοδοι σε σύγκριση με άλλες σε πόρους, αφού πρέπει να πληρούνται αρκετές προϋποθέσεις για να εφαρμοστούν. Στα μειονεκτήματά τους καταλογίζονται ότι δεν βρίσκουν εν γένει τα προβλήματα ασυνέπειας της διασύνδεσης (αφού οι χρήστες δεν μπορούν να τα καταδείξουν) όπως και τα μικρής σοβαρότητας προβλήματα (που οι χρήστες τα ξεπερνούν και τα αγνοούν στη συνέχεια).

☞ **Ο Μάγος του Οζ (Wizard of Oz).** (Nielsen, 1993b)

Ένα μέλος της σχεδιαστικής ομάδας είναι «κρυμμένο» και αντιδρά για λογαριασμό του συστήματος στις ενέργειες του χρήστη, κυρίως σε αυτές που δεν έχουν ακόμα ενσωματωθεί στο σύστημα.

Υπέρ	Κατά
<ul style="list-style-type: none"> Χρησιμοποιείται ακόμα και αν δεν είναι έτοιμη μια «έκδοση εργασίας – working version» του συστήματος. Δίνει αρκετή ανάδραση για το τι περιμένει ο χρήστης αρκετά νωρίς 	<ul style="list-style-type: none"> Δύσκολο γι' αυτόν που θα κάνει τον «μάγο» Δεν είναι παντα εφικτό να αντιδράσει κανείς στις ενέργειες του χρήστη ή μπορεί να είναι έξω από τις δυνατότητες του συστήματος.

Συνιστάται όταν θέλουμε γρήγορη ανάδραση (feedback) χωρίς να διαθέτουμε ακόμα έτοιμο πρωτότυπο.

☞ **Φωνητική Σκέψη (Thinking Aloud Protocol).**

Ζητάμε από τον χρήστη να εκφράζει δυνατά τη σκέψη του κάθε στιγμή σε σχέση με αυτά που κάνει. Η αρχική ιδέα εφαρμόζεται παραδοσιακά σε ψυχολογικές ερευνητικές μεθοδολογίες (Ericsson & Simon, 1984), όμως η προσαρμογή της στην αξιολόγηση διασυνδέσεων καθώς και κρίσιμες λεπτομέρειες για την οργάνωση και τη διεξαγωγή της συνεδρίας περιγράφονται στο Lewis (1982).

Υπέρ	Κατά
<ul style="list-style-type: none"> Αποκαλύπτει πληροφορίες που δεν θα εμφανίζονταν αλλιώς, πχ. τις παρανοήσεις των χρηστών Εμφανίζει το γνωστικό κόσμο του χρήστη Αποκαλύπτει και το τι και το γιατί συνέβη Επιτρέπει επέμβαση του παρατηρητή και δρομολόγηση της αξιολόγησης Χρειάζονται σχετικά λίγοι χρήστες Τα σχόλια των χρηστών δίνουν πολλές υποδείξεις 	<ul style="list-style-type: none"> Ο χρήστης αφοσιώνεται και παύει να μιλάει ή ο χρήστης, προσπαθώντας να συνεχίσει να μιλάει, κάνει λάθη στη χρήση που δεν θα έκανε κανονικά. Τα φαινόμενα αυτά είναι πιο έντονα σε έμπειρους χρήστες. Πολλές φορές οι χρήστες δίνουν λάθος εξηγήσεις γι' αυτά που κάνουν Τα σχόλια των χρηστών μπορούν να παραπλανήσουν Χρειάζεται έμπειρος παρατηρητής Δεν επιτρέπει μετρήσεις επίδοσης Σχετικά δύσκολη επεξεργασία αποτελεσμάτων

Συνιστάται όταν θέλουμε να κατανοήσουμε το **γιατί** ο χρήστης κάνει αυτά που κάνει.

Σημείωση: Η μέθοδος μπορεί να χρησιμοποιηθεί και για το σχεδιασμό του interface, όπως για παράδειγμα στο Maulsby et al. (1993) με σκοπό να εμφανίσουμε τους γνωστικούς δρόμους που τείνει να ακολουθήσει ο χρήστης.

☞ **Εποικοδομητική Διάδραση (Constructive Interaction)** (O'Malley et al., 1984) ή **Συναποκαλυπτική Μάθηση (Codiscovery Learning)** (Kennedy, 1989)

Δύο χρήστες – αξιολογητές χρησιμοποιούν μαζί το σύστημα, συζητώντας μεταξύ τους ό,τι συναντάνε.

Υπέρ	Κατά
<ul style="list-style-type: none"> Η αξιολόγηση είναι πιο φυσική Μπορεί να αποκαλύψει προβλήματα συνεργασίας σε διασυνδέσεις που εμπλέκουν ομάδες χρηστών 	<ul style="list-style-type: none"> Οι χρήστες μπορεί να έχουν διαφορετικές στρατηγικές χειρισμού υπολογιστών ή διαφορετικό τρόπο εκμάθησης της διασύνδεσης Χρειάζεται διπλάσιους χρήστες που μπορεί να σημαίνει διπλάσιο κόστος

Συνιστάται σε διασυνδέσεις που στοχεύουν σε παιδιά, αφού αυτά, σαν χρήστες – αξιολογητές δεν μπορούν να ακολουθήσουν εύκολα τις οδηγίες διεξαγωγής της αξιολόγησης.

☞ **Ανασκοπιστικός Έλεγχος (Retrospective Testing)** (Hewett & Scott, 1987)

Βιντεοσκοπούμε τη συνεδρία και ξαναβλέπουμε τη βιντεολήψη μαζί με το χρήστη, ρωτώντας τον για διευκρινίσεις.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Ο χρήστης έχει περισσότερη άνεση για σχόλια • Ο άγων την αξιολόγηση μπορεί να επιμένει σε κάποια σημεία 	<ul style="list-style-type: none"> • Κάθε συνεδρία χρειάζεται τουλάχιστον τον διπλάσιο χρόνο, άρα πόρους • Δεν είναι πάντα σίγουρο ότι οι χρήστες θυμούνται γιατί έκαναν κάτι και μπορεί να δώσουν αναξιόπιστες εξηγήσεις

Συνιστάται όταν οι χρήστες – αξιολογητές είναι δυσέυρετοι, οπότε πρέπει να «στοραγγίξουμε» όσους βρούμε.

☞ **Η Μέθοδος του «Προπονητή» (Coaching Method)** (Mack & Burdett, 1992)

Στη μέθοδο αυτή, σε αντίθεση με τις περισσότερες άλλες, ο άγων την αξιολόγηση αλληλεπιδρά με το χρήστη – αξιολογητή και προσπαθεί να τον καθοδηγήσει στη σωστή κατεύθυνση.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Ο χρήστης μπορεί να ρωτάει για τη διασύνδεση • Μπορεί να χρησιμοποιηθεί σαν τρίτος «προπονητής» ένας έμπειρος χρήστης, οπότε ο άγων παρατηρεί την αλληλεπίδραση • Πιο φυσική μέθοδος 	<ul style="list-style-type: none"> • Απαιτεί επιδέξιο άγοντα την αξιολόγηση • Συνήθως εστιάζεται μόνο σε άπειρους χρήστες • Χρονοβόρο, αφού κάθε διαδικασία επιμηκύνεται ανάλογα με την διάδραση των εμπλεκομένων

Συνιστάται όταν δεν έχουμε εύκαιρους πολλούς χρήστες ή έχουμε δύσκολη διασύνδεση.

Σημείωση: Η μέθοδος αποδείχθηκε πολύτιμη σε ιάπωνες χρήστες (Kato, 1986), πολλοί από τους οποίους, λόγω πολιτισμικών περιορισμών, δεν μπορούσαν να εκφράσουν αλλιώς τη δυσαρέσκειά τους σε συγκεκριμένες διεργασίες της διασύνδεσης.

☞ **Συμμετοχική Δομημένη Παρατήρηση**

Οι χρήστες-αξιολογητές εργάζονται με το υπό αξιολόγηση προϊόν, ακολουθώντας επακριβώς μια διαγεγραμμένη πορεία διεργασιών και ενεργειών (task script) (Kantner & Rosenbaum, 1997) υπό ελεγχόμενες συνθήκες. Ένας αξιολογητής άγει την αξιολόγηση, ενώ ένας άλλος σημειώνει.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Δίνει στατιστικά επεξεργάσιμα αποτελέσματα (μετρήσιμα) • Απαντάει στην ερώτηση «γιατί» • Απαντάει σε ερωτήσεις ειδικές του συγκεκριμένου περιβάλλοντος 	<ul style="list-style-type: none"> • Χρονοβόρο • Κοστίζει • Χάνει την εικόνα του όλου • Δεν αξιολογεί συνολικά τη διασύνδεση

Συνιστάται όταν έχουμε διαθέσιμους χρήστες που μπορούν να απαντήσουν έγκυρα

☛ Συμμετοχική Ελεύθερη Παρατήρηση (Diaper, 1989)

Οι χρήστες-αξιολογητές εργάζονται ελεύθερα με το υπό αξιολόγηση προϊόν, ακολουθώντας μια πραγματική πορεία για την ολοκλήρωση των στόχων τους σε όσο το δυνατό πιο πραγματικό εργασιακό περιβάλλον. Ο άγων την αξιολόγηση περνάει σχεδόν απαρατήρητος.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Οι χρήστες εργάζονται σε πολύ φυσικές συνθήκες, αποκαλύπτοντας τις πραγματικές διεργασίες που χρησιμοποιούν • Συχνά ανακαλύπτεται ότι οι χρήστες χρησιμοποιούν το σύστημα διαφορετικά από ότι πρόβλεψαν οι σχεδιαστές 	<ul style="list-style-type: none"> • Ο άγων δεν αντιδρά σχεδόν καθόλου, συνεπώς μένουν πολλά ερωτήματα αναπάντητα • Χρειάζεται έμπειρος άγων για να διακρίνει τα προβλήματα • Δεν έχουμε έλεγχο στη ροή της συνεδρίας

Συνιστάται όταν θέλουμε δεδομένα από πραγματική χρήση.

☛ Μη Συμμετοχική Παρατήρηση

Οι χρήστες-αξιολογητές εργάζονται με το υπό αξιολόγηση προϊόν, όμως δεν γνωρίζουν ότι παρατηρούνται. Γενικά αυτός ο τρόπος αξιολόγησης δεν συνιστάται για ηθικούς λόγους, όμως ίσως υπάρχουν περιπτώσεις όπου αυτός είναι ο μόνος τρόπος (πχ. σε ένα πληροφοριακό κιόσκι – info kiosk). Σ' αυτή την περίπτωση τα αποτελέσματα καταγράφονται συνήθως με κάμερες ή λογισμικό καταγραφής (logging software).

Υπέρ	Κατά
<ul style="list-style-type: none"> • Οι χρήστες αλληλεπιδρούν με το σύστημα σε τελείως φυσικές συνθήκες 	<ul style="list-style-type: none"> • Δεν ξέρουμε σχεδόν ποτέ το λόγο που οι χρήστες αντιδρούν κατά συγκεκριμένο τρόπο

Συνιστάται όταν δεν υπάρχει άλλος τρόπος για αξιολόγηση με χρήστες.

Καταμετροητικές Εμπειρικές (survey empirical evaluation)

Είναι διαμορφωτικές/αποτιμητικές – ποιοτικές/ποσοτικές.

Χρησιμοποιούνται συνεντεύξεις ή ερωτηματολόγια για να εξαχθούν οι υποκειμενικές γνώμες των χρηστών για το πως αντιλαμβάνονται τη διασύνδεση. Είναι συνεπώς έμμεσοι μέθοδοι καταγραφής, αφού δεν συλλέγονται δεδομένα από αυτή καθ' αυτή τη χρήση του συστήματος, αλλά από τη γνώμη των χρηστών γι' αυτό. Οι συνεντεύξεις χρησιμοποιούνται με ένα περιορισμένο γκρουπ χρηστών και μπορεί να είναι χρονοβόρες. Τα ερωτηματολόγια μπορούν να χρησιμοποιηθούν με μεγάλα γκρουπ χρηστών και είναι ευκολότερα επεξεργάσιμα. Πλεονέκτημα όλων αυτών των μεθόδων είναι ότι μπορούν να καταγράψουν την παράμετρο της υποκειμενικής ικανοποίησης του χρήστη που δεν μπορεί να καταγραφεί από τις άλλες μεθόδους και ως προς αυτή την παράμετρο συνιστούν άμεση μέθοδο αξιολόγησης (Nielsen, 1993b).

☛ Ερωτηματολόγιο (Questionnaire).

Οι χρήστες απαντούν εγγράφως σε ερωτήσεις σχετικές με τη διασύνδεση.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Μπορεί να επιδοθεί σε μια μεγαλύτερη ομάδα ατόμων • Χρειάζεται λιγότερο χρόνο απάντησης, αφού στον ίδιο χρόνο μπορούν να 	<ul style="list-style-type: none"> • Γενικά, το ερωτηματολόγιο είναι λιγότερο ευέλικτο, αφού οι ερωτήσεις είναι συγκεκριμένες από την αρχή. Χρειάζεται πιλοτική φάση για την αποφυγή

<p>απαντήσουν πολλοί χρήστες</p> <ul style="list-style-type: none"> • Τα στοιχεία του μπορούν να αναλυθούν εύκολα και να επεξεργαστούν στατιστικά • Μπορεί να ξαναχρησιμοποιηθεί 	<p>παρεξηγήσεων.</p> <ul style="list-style-type: none"> • Οι απόψεις των χρηστών μπορεί να μην είναι έγκυρες ή να είναι προκατειλημμένες • Ευαίσθητο στο μέγεθος και στη δομή του
--	---

Συνιστάται όταν έχουμε μεγάλο αριθμό χρηστών – αξιολογητών.

☞ **Ανοικτή Συνέντευξη (Open Interview)**

Μια ανοικτή συνέντευξη αρχίζει συνήθως με απλές και γενικές ερωτήσεις, οι οποίες στη συνέχεια εξειδικεύονται περισσότερο. Στο είδος αυτό της συνέντευξης δεν είναι καθορισμένες επακριβώς οι ερωτήσεις που θα τεθούν στο χρήστη, παρά μόνο οι γενικοί άξονες που πρέπει να ακολουθηθούν για την απάντηση των ερωτημάτων της αξιολόγησης.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Ευελιξία: Το επίπεδο των ερωτήσεων μπορεί να ποικίλλει, για να δοθεί μεγαλύτερο βάρος σε κάποιο ζήτημα που ίσως προκύψει κατά τη διάρκεια της συνέντευξης (διαδοχικές ερωτήσεις) • Αποκαλύπτει συχνά πράγματα που δεν περιμέναμε • Δύσκολη επεξεργασία • Κατάλληλη για «διερευνητική εμπάθунση», με άλλα λόγια αν δεν ξέρουμε τι ακριβώς ψάχνουμε. 	<ul style="list-style-type: none"> • Η πιθανή υποκειμενικότητα-μεροληψία τόσο του ερευνητή-αξιολογητή όσο και του ερωτώμενου. • Σχεδόν αδύνατη η στατιστική επεξεργασία • Δυσκολία στην ταξινόμηση και επεξεργασία των απόψεων • Χρονοβόρα μέθοδος • Χρειάζεται έμπειρος άγων τη συνέντευξη

Συνιστάται όταν μας ενδιαφέρουν μόνο ποιοτικά αποτελέσματα

☞ **Ημιδομημένη Συνέντευξη (Semi-Structured Interview)**

Η συνέντευξη αυτή ακολουθεί μια εκ των προτέρω προδιαγεγραμμένη πορεία μόνο ως προς το βασικό κορμό ερωτήσεων που πρέπει να απαντηθούν. Ακολουθώντας καλούνται οι συμμετέχοντες να εκφράσουν ελεύθερα τη γνώμη τους.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Διατηρεί το σκοπό της αξιολόγησης, αλλά δίνει ελευθερία στο χρήστη να εκφραστεί 	<ul style="list-style-type: none"> • Χρονοβόρα μέθοδος • Χρειάζεται έμπειρος άγων τη συνέντευξη • Σχεδόν αδύνατη η στατιστική επεξεργασία

Συνιστάται όταν έχουμε «δύσκολες» ομάδες χρηστών, πχ. αναλφάβητους

☞ **Δομημένη Συνέντευξη (Structured Interview)**

Η συνέντευξη αυτή ακολουθεί μια εκ των προτέρω προδιαγεγραμμένη πορεία. Θα πρέπει δηλαδή να έχει καθοριστεί από πριν ένας βασικός κορμός ερωτήσεων και οι περισσότερες από τις υποερωτήσεις που πρέπει να απαντηθούν από το χρήστη-αξιολογητή, καθώς και η δομή της όλης διαδικασίας, ακόμα και της πρόβλεψης πιθανών απαντήσεων, ακόμα και το πώς θα μπορούσε ο ερευνητής-αξιολογητής να τις χειριστεί κατά τη διεξαγωγή της συνέντευξης.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Μπορεί να μετετραπεί εύκολα σε στατιστικά επεξεργάσιμη μορφή 	<ul style="list-style-type: none"> • Θέλει αρκετή προετοιμασία • Διευκρινίζει μόνο ό,τι προβλέφθηκε από πριν. • Χρονοβόρα μέθοδος

Συνιστάται όταν επιζητούμε την απάντηση συγκεκριμένων ερωτημάτων

☞ **Ομαδική Συνέντευξη (Group Interview)**

Οι ομαδικές συνεντεύξεις αποτελούν επίσης μια σημαντική κατηγορία, εξαιτίας της έμφασης που δίνεται στην κριτική ανάλυση των απόψεων και δεδομένων από τους συμμετέχοντες.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Τόσο οι ομοφωνίες όσο και οι διαφωνίες μιας μεγαλύτερης ομάδας χρηστών γίνονται αντιληπτές • Παράγει συγκεντρωτικά αποτελέσματα • Μπορούν να ανακαλυφθούν προβλήματα που οι διακριτοί αξιολογητές δεν καταδεικνύουν 	<ul style="list-style-type: none"> • Χαλαρή δομή, μπορεί να «ξεστρατίσει» • Δύσκολη η καταγραφή μεμονωμένων απόψεων • Δύσκολη και η ακριβής καταγραφή των συνολικών τάσεων. Αν χρησιμοποιηθεί βιντεοσκόπηση, μεγάλος χρόνος επεξεργασίας • Στην πράξη παρουσιάζεται πρόβλημα συνάθροισης των αξιολογητών • Χρειάζεται έμπειρος άγων τη συνέντευξη

Συνιστάται όταν θέλουμε να αποτιμήσουμε τη γενική εντύπωση και τη γενική αποδοχή του προϊόντος

☞ **Λογισμικό Καταγραφής (Logging Software)**

Το υπολογιστικό σύστημα καταγράφει τις κινήσεις του χρήστη και καταμετρεί τους χρόνους.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Πραγματικού χρόνου καταγραφή • Αναλυτική καταγραφή όλων των κινήσεων, άρα μεγάλη ποσότητα δεδομένων • Ποσοτικά δεδομένα • Καταδεικνύει τα λάθη των χρηστών • Μπορεί να διαρκέσει όσο θέλουμε και να συλλέξει δεδομένα από πολλούς χρήστες • Μπορεί να καταγράψει δεδομένα που ξεφεύγουν από τις άλλες μεθόδους, πχ. άστοχα κλικ του ποντικιού ή ορθογραφικά λάθη του χρήστη. 	<ul style="list-style-type: none"> • Δεν μπορεί να απαντήσει το «γιατί» • Δεν καταγράφει ενδιάμεσες κινήσεις, παρά μόνο τα κλικ ή συγκεκριμένες ενέργειες • Μεγάλος όγκος δεδομένων για επεξεργασία • Δεν καταδεικνύει τον ψυχικό και τον γνωστικό κόσμο του χρήστη • Στην πράξη δεν είναι δυνατό να αξιολογηθούν όλες οι πλευρές και οι δυνατότητες της διασύνδεσης • Δυσκολία προγραμματισμού και υλοποίησης • Παραβίαση της ιδιωτικότητας (privacy) του χρήστη

Συνιστάται όταν ζητάμε πληροφορίες σχετικές με την κατασκευή ως προς ευχρηστία του λογισμικού (usability engineering). Για παράδειγμα οι Senay & Stabler (1987) κατέγραψαν 52.576 αιτήσεις βοήθειας (help) χρηστών κάποιου λογισμικού και βρήκαν ότι το 92% αυτών των αιτήσεων αφορούσε το 10% των συνολικών οθονών βοήθειας που παρείχε το πρόγραμμα.

☞ **Ομάδες Εστίασης (Focus Groups)** (Caplan, 1990; Goldman & McDonald, 1987; Greenbaum, 1988; Greenbaum, 1993; O'Donell et al., 1991)

Η προσέγγιση αυτή δεν είναι στην ουσία αξιολογητική μέθοδος, αλλά μέθοδος του usability engineering, σύμφωνα όμως με τους O'Donell et al. (1991) μπορεί να χρησιμοποιηθεί και σαν αξιολογητική. Συγκεντρώνουμε έξι με εννέα χρήστες οι οποίοι υπό την καθοδήγηση ενός συντονιστή συζητούν για δύο ώρες θέματα της διασύνδεσης. Οι χρήστες αισθάνονται

ελεύθεροι, όμως ο συντονιστής φροντίζει να διατηρείται η εστίαση και να συζητηθούν όλα τα προγραμματισθέντα θέματα.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Εμφανίζει όλες τις ιδέες των χρηστών • Εμφανίζει τις διεργασίες των χρηστών (user tasks) • Εμφανίζει την ομαδική δυναμική και την οργάνωση του συνολικού γκρουπ χρηστών που απευθύνεται το σύστημα 	<ul style="list-style-type: none"> • Χρειάζεται η προετοιμασία λίστας θεμάτων και στόχων • Χρειάζεται έμπειρος άγων τη συζήτηση • Δυσκολία επεξεργασίας των δεδομένων • Κάποιοι από τους χρήστες μπορεί να κυριαρχούν (dominate) • Δυσκολία στη συνάθροιση των χρηστών • Τα αποτελέσματα μπορεί να έχουν χαμηλή εφαρμοσιμότητα

Συνιστάται πολύ νωρίς στο σχεδιαστικό κύκλο για να βγούν οι προδιαγραφές, ή πολύ αργά μετά την κυκλοφορία του προϊόντος για να συλλεχθούν οι αποτιμήσεις.

☞ **Ανάδραση Χρηστών (User Feedback)** (Nielsen, 1993b, σελ. 221)

Ταχυδρομικά στέλνονται ερωτηματολόγια στους χρήστες, ή περιλαμβάνονται στη συσκευασία του προϊόντος. Προφανώς είναι αποτιμητική μέθοδος που μπορεί να εφαρμοστεί μόνο μετά την επίσημη κυκλοφορία του προϊόντος.

Υπέρ	Κατά
<ul style="list-style-type: none"> • Είναι μια διαδικασία διαρκείας με συνεχή ανάδραση και εμπλέκει μεγάλους αριθμούς χρηστών • Καταδεικνύει τόσο τις τωρινές τους ανησυχίες, όσο και τη μεταβολή των στάσεών τους με το χρόνο 	<ul style="list-style-type: none"> • Χρειάζεται ειδική οργάνωση για την επεξεργασία των δεδομένων και τη διαχείριση των χρηστών-αξιολογητών • Μπορεί να μην είναι αντιπροσωπευτικό, αφού απαντούν συνήθως οι «φλύαροι» και οι δυσάρεστοι.

Συνιστάται όταν σκοπεύουμε να βελτιώσουμε το προϊόν με επόμενη έκδοση.

1.9. Προβλήματα των αξιολογήσεων

Υπάρχουν κάποια προβλήματα που τείνουν να παρουσιάζουν γενικά όλες οι μεθοδολογίες αξιολογήσεων. Η λίστα αυτών είναι τόσο μεγάλη που πρακτικά δεν μπορεί κανείς να τα αποφύγει όλα. Μπορούμε ενδεικτικά να αναφέρουμε κάποιες συνηθισμένες πηγές προβλημάτων, όπως:

- Το χρόνο
- Το κόστος
- Τη μεροληψία
- Την υποκειμενικότητα
- Την ιδεολογική φόρτιση των αξιολογητών ή και των φορέων αξιολόγησης

όσον αφορά

- Την επιλογή των κύριων ερωτημάτων
- Τον τρόπο συλλογής
- Επεξεργασίας
- Ανάλυσης
- Ερμηνείας των πορισμάτων μιας αξιολόγησης,

όπως επίσης και μη προβλέψιμα προβλήματα που πολλές φορές αποβαίνουν καταστροφικά, όπως:

- Εσφαλμένες επιλογές τόσο κατά τη διαδικασία της αξιολόγησης,
- Όσο και εξαιτίας της έλλειψης επαρκών ικανοτήτων αυτών που διεξάγουν την αξιολόγηση
- Ή αντιδράσεων τρίτων στο έργο της αξιολόγησης,
- Ή μη συνεργασίας των συμμετεχόντων στο έργο της αξιολόγησης.

Προβλήματα μπορούν επίσης να υπάρξουν και από τις τυχόν

- Παραλείψεις
- Σφάλματα
- Παρεμβάσεις
- Μεροληψίες

που μπορεί να έχουν γίνει στα δευτερογενή δεδομένα.

Υπάρχουν μερικοί βασικοί τρόποι περιορισμού των επιδράσεων αυτών των προβλημάτων. Ο Μακράκης (1999) για παράδειγμα συνιστά:

- Τη χρησιμοποίηση έμπειρων αξιολογητών.
- Τη χρησιμοποίηση περισσότερων του ενός αξιολογητών.
- Τη χρησιμοποίηση εξωτερικών αξιολογητών που δεν πρέπει να εμπλέκονται στενά με αυτό που αξιολογείται.
- Τη χρήση πολλαπλών μεθόδων και τεχνικών αξιολόγησης.
- Το σωστό σχεδιασμό και τη συστηματική οργάνωση της αξιολόγησης.
- Την παραδοχή και εξωτερίκευση των αξιών που καθορίζουν την αξιολόγηση.

Όπως έχουμε όμως ήδη συζητήσει, αυτές οι συστάσεις, όσο και να φαίνονται πολλά υποσχόμενες, μπορούν να υλοποιηθούν μόνο όπου το επιτρέπει η υπό αξιολόγηση οντότητα και η επιλεγείσα μέθοδος αξιολόγησης.

Εκτός όμως από τα προαναφερθέντα βασικά προβλήματα, υπάρχουν και κατά τη διαδικασία της αξιολόγησης πολλές λεπτομέρειες που όσο και να προσπαθήσει ο ερευνητής-αξιολογητής ίσως δεν μπορεί να αποφύγει και οι οποίες μπορούν να προκαλέσουν προβλήματα στην επεξεργασία των αποτελεσμάτων. Οι Lewis & Rieman (1994) αναφέρουν κάποια από αυτά τα προβλήματα, όπως:

Τί κάνουμε αν κάποιος αξιολογητής δεν ολοκληρώνει μια διεργασία; Είναι πολύ πιθανό ότι κάποιοι αξιολογητές δεν θα μπορέσουν να ολοκληρώσουν τις διεργασίες ή δεν θα τις ολοκληρώσουν στον προβλεπόμενο χρόνο, ή δεν θα μπορούν χωρίς τη βοήθειά μας. Δεν υπάρχει σαφής λύση γι' αυτό το πρόβλημα. Μια προσέγγιση είναι να θεωρήσουμε ένα πολύ μεγάλο χρόνο ολοκλήρωσης ή ένα πολύ μεγάλο αριθμό επιτρεπτών λαθών γι' αυτά τα άτομα. Προφανώς τα αποτελέσματα της ανάλυσης μας θα είναι τότε επίσης με ένα μεγαλύτερο βαθμό ανοχής.

Κρατώντας χαμηλά τη διαφοροποίηση. Για να μπορέσουμε να έχουμε εύκολα επεξεργάσιμα δεδομένα είναι επίσης βασικό να έχουμε αποτελέσματα με μικρές διαφοροποιήσεις. Υπάρχουν τεχνικές για να κρατήσουμε χαμηλά τις διαφοροποιήσεις αυτές, όμως επιστάται η προσοχή ότι στην πραγματική χρήση τα πράγματα δεν θα είναι έτσι. Η κυριώτερη πηγή για πολύ διαφορετικά αποτελέσματα είναι οι πολύ διαφορετικοί αξιολογητές. Αν είναι μάλιστα πολύ διαφορετικοί τόσο στην εμπειρία τους με τους υπολογιστές, όσο και στην αξιολογούμενη διεργασία, τότε είναι αναμενόμενη μια μεγάλη διαφοροποίηση στους χρόνους ολοκλήρωσης και τους ρυθμούς σφαλμάτων. Εδώ βοηθάει αποτελεσματικά μια πιο προσεκτικά οργανωμένη επεξήγηση της αξιολόγησης και των διεργασιών που πρέπει να εκτελέσουν. Αν οι χρήστες δεν καταλαβαίνουν τι κάνουν, τότε αυτό πρέπει να διαλευκανθεί, αλλιώς η διαφοροποίηση είναι δεδομένη.

Μπορεί όμως να είμαστε και εμείς πηγή διαφοροποιήσεων, κυρίως αν αλλάζουμε τον τρόπο που διενεργούμε την αξιολόγηση. Αν για παράδειγμα βοηθάμε κάποιους αξιολογητές περισσότερο από άλλους ή αλλάζουμε τη σειρά των διεργασιών. Αυτό βέβαια είναι ακόμα ένα επιχείρημα για την προσεκτικότερη προετοιμασία της αξιολόγησης.

Ζητώντας διευκρινίσεις από τους αξιολογητές. Γενικά δεν συνιστάται να ρωτάμε ή να ενοχλούμε τους αξιολογητές κατά τη διάρκεια της αξιολόγησης. Όμως τι γίνεται στην περίπτωση που κάποιος αξιολογητής έχει τελειώσει μια διεργασία και μπορεί να μας εξηγήσει κάτι; Υπάρχει και αυτό σαν μέθοδος αξιολόγησης (και θα αναφερθεί παρακάτω στην ταξινόμηση), όμως οι Lewis & Rieman (1994) υποστηρίζουν ότι η προσέγγιση αυτή δεν είναι αξιόπιστη. Αναφέρουν ότι από έρευνές τους οι χρήστες (αναφέρονται σε αξιολογήσεις με χρήστες) αμέσως μετά από την αξιολόγηση και σχετικά με τα προβλήματα που αντιμετώπισαν δεν θυμόνταν και πολλά πράγματα σχετικά με αυτά που έκαναν, αλλά ούτε και το γιατί τα έκαναν και ακόμα περισσότερο στην περίπτωση που μπόρεσαν να ξεπεράσουν αυτά τα προβλήματα, ακόμα και με δυσκολία. Οι ερευνητές αυτοί υποστηρίζουν ότι σ' αυτή την περίπτωση ο χρήστης-αξιολογητής θυμάται ίσως τη λύση, αλλά όχι το πρόβλημα.

Η λύση σ' αυτό το πρόβλημα φαίνεται να είναι η βιντεοσκόπηση της συνεδρίας (Nielsen, 1993b). Έτσι μπορούμε να ξαναδείξουμε στους αξιολογητές το συγκεκριμένο σημείο και να τους ρωτήσουμε γιατί αντέδρασαν έτσι. Εδώ οι γνώμες διίστανται. Ο Nielsen (1993b) υποστηρίζει ότι έτσι μπορεί να λυθεί το πρόβλημα, ενώ οι Lewis & Rieman (1994) προειδοποιούν να μην περιμένουμε πολλά. Υποστηρίζουν ότι οι αξιολογητές θα υποθέσουν μάλλον το ίδιο που σκεφτόμαστε και εμείς, σχετικά με το γιατί αντέδρασαν κατά αυτόν τον τρόπο τη συγκεκριμένη στιγμή.

Άλλος τρόπος διευκρινίσεων που φαίνεται να αποδίδει είναι το να ρωτάμε σχόλια για συγκεκριμένες δυνατότητες (features) της διασύνδεσης. Συνήθως υπάρχουν υποδείξεις και αντιδράσεις, θετικές ή αρνητικές, που δεν θα μπορούσαν με άλλο τρόπο να αποτυπωθούν στα δεδομένα μας. Μάλιστα αυτή η μέθοδος θα δουλέψει ακόμα καλύτερα, αν με κάποιο τρόπο μπορέσουμε να αναπαραστήσουμε στον αξιολογητή την αλληλουχία των οθονών που διέτρεξε στη διάρκεια της αξιολόγησης. (Lewis & Rieman, 1994)

1.10. Συζήτηση

Στα επόμενα συζητούνται κάποια θέματα, ανεξάρτητα το ένα από το άλλο, που μπορούν να προκύψουν κατά τις διεξαγωγές των αξιολογήσεων ανάλογα με την ακολουθούμενη μεθοδολογία και όλα μαζί συνθέτουν την πολύπλοκη εικόνα αυτού του πεδίου της επικοινωνίας του ανθρώπου με τον υπολογιστή.

1. Η Αξιοπιστία των Μεθοδολογιών Αξιολόγησης

Το θέμα αυτό είναι το θέμα με τη μεγαλύτερη σημασία όταν συζητάει κανείς για αξιολόγηση. Και η καλύτερη μέθοδος είναι πρακτικά άχρηστη όταν δεν παράγει αξιόπιστα αποτελέσματα. Η παρούσα διατριβή αφιερώνει ένα κεφάλαιο στο ζήτημα αυτό, το «*Κεφάλαιο 3: Η Παράμετρος της Αξιοπιστίας στην Αξιολόγηση – Η Συνδυαστική Αξιολόγηση*», όπου συζητείται εκτενώς το θέμα αυτό. Συνεπώς ο βιαστικός αναγνώστης παραπέμπεται εκεί, ενώ μια πρώτη γεύση του προβληματισμού δίνεται στην αμέσως παρακάτω παράγραφο.

2. Αξιολογήσεις βασισμένες σε ειδικούς (expert-based evaluation methodologies) και αξιολογήσεις με χρήστες (user-based evaluation methodologies)

Σύμφωνα με τους Lewis & Rieman (1994) «δεν μπορείς να πείς πόσο καλό είναι το interface αν δεν βρεις ανθρώπους να το χρησιμοποιήσουν». Αυτή η φράση εκφράζει συμπεπυγμένα την πλατιά διαδεδομένη άποψη σύσσωμης της βιβλιογραφίας ότι οι αξιολογήσεις με χρήστες είναι απαραίτητες και τελικά οι περισσότερες αποδοτικές για την αποτίμηση της ευχρηστίας μιας διασύνδεσης. Γιατί λοιπόν δεν χρησιμοποιούμε αποκλειστικά εμπειρικές αξιολογήσεις, με χρήστες δηλαδή, αλλά ερευνάμε και άλλες προσεγγίσεις; Όπως θα δούμε στη συνέχεια, η αποδοτικότητα των εμπειρικών αξιολογήσεων μετριάζεται ισχυρά από τους πόρους που καταναλώνουν και από κάποια μειονεκτήματα που παρουσιάζουν, ενώ παράλληλα οι αξιολογήσεις βασισμένες σε ειδικούς έχουν πλέον ωριμάσει αρκετά ώστε να αποτελούν καλή εναλλακτική λύση.

Οι Lewis & Rieman (1994) δίνουν τρία πράγματα που χρειάζονται για μια εμπειρική αξιολόγηση:

- Ανθρώποι και κατά προτίμηση πραγματικοί χρήστες σε πραγματικές συνθήκες
- Κάποιες διεργασίες για να εκτελέσουν και
- Μια έκδοση του συστήματος για να δουλέψουν

Στο σημείο αυτό μπορεί ήδη να έχουμε το πρώτο κώλυμα εφαρμογής εμπειρικής αξιολόγησης, καθώς τα στοιχεία αυτά είναι και τα τρία συγχρόνως απαραίτητα. Δεν είναι όμως πάντα δυνατόν να συντονιστούν, όπως καταλαβαίνει εύκολα κανείς.

Από την άλλη πλευρά οι αξιολογήσεις οι βασισμένες σε ειδικούς αποτελούν σύμφωνα με τον Reeves (1993) ίσως την συνηθέστερα εφαρμοζόμενη στρατηγική αξιολόγησης. Γιατί αυτό; Οι μεθοδολογίες αξιολόγησης με ειδικούς έχουν ένα κρίσιμο πλεονέκτημα που τις κάνει τόσο προσιτές σε σχέση με τις αξιολογήσεις με πραγματικούς χρήστες: είναι εν γένει πολύ ευκολότερο και πολύ οικονομικότερο να βρεί κανείς ειδικούς παρά χρήστες πρόθυμους να αξιολογήσουν ένα προϊόν. Η βασική ιδέα είναι ότι ειδικοί από διάφορα γνωστικά πεδία, αλλά οπωσδήποτε και με γνώσεις από το HCI, καλούνται να κρίνουν τη διασύνδεση από τη δική του οπτική γωνία ο καθένας. Τίθεται εδώ ως προϋπόθεση ότι η εμπειρία τους σαν ειδικοί θα τους επιτρέψει να δουν τη διασύνδεση «με τα μάτια του χρήστη» και να εντοπίσουν πιθανά προβλήματα που θα αντιμετωπίσει αυτός κατά τη χρήση του συστήματος. Ένα από τα ισχυρά πλεονεκτήματα των μεθόδων αυτών είναι ότι μπορούν να εφαρμοστούν ακόμα και πολύ νωρίς στο σχεδιαστικό κύκλο, ακόμα και σε σχέδια στο χαρτί (paper mock-ups). Εκεί η εμπειρία του ειδικού του επιτρέπει να κατανοεί την λειτουργία της εφαρμογής και δεν χρειάζεται να έχει την πλήρη εικόνα του ολοκληρωμένου προϊόντος, αλλά μια

«πρώτη εικόνα» με τα βασικά χαρακτηριστικά αρκεί. Δεν ισχύει προφανώς αυτό για τον απλό χρήστη και οι μεθοδολογίες οι βασισμένες σε χρήστες αρχίζουν να εφαρμόζονται εν γένει μόνο αφού το προϊόν έχει φτάσει σε σημαντικό βαθμό ολοκλήρωσης.

Επίσης πολύ βασικό για τις αξιολογήσεις με πραγματικούς χρήστες είναι το να βρούμε αντιπροσωπευτικούς χρήστες, σαν αυτούς που θα χρησιμοποιήσουν τελικά στην πράξη το προϊόν. Όπως λένε χαρακτηριστικά οι Lewis & Rieman (1994) «αν δεν μπορείτε να βρείτε κανονικούς χρήστες για να αξιολογήσετε το προϊόν σας, τότε τι σας κάνει να πιστεύετε ότι θα βρείτε χρήστες να το χρησιμοποιήσουν;» Συνεπώς εδώ έχει εφαρμογή πάνω από όλα η προτροπή του Nielsen (1993b) «Γνώριζε τους χρήστες σου! – know thy users!», όπως και η προτροπή των Lewis & Rieman (1994) ότι το πρώτο βήμα για τον χρηστοκεντρικό σχεδιασμό μιας διασύνδεσης (user-centered interface design) είναι να επισκεφθεί η ομάδα σχεδιασμού τους πραγματικούς χρήστες στον πραγματικό χώρο εργασίας τους και να καταγράψει τις ακριβείς διεργασίες (tasks) που αυτοί χρησιμοποιούν στη διάρκεια της δουλειάς τους. Με το ίδιο σκεπτικό, η αξιολόγηση με χρήστες πρέπει να προσομοιώνει όσο το δυνατό καλύτερα τις πραγματικές συνθήκες χρήσης του προϊόντος, με αντιπροσωπευτικό δείγμα χρηστών για να έχει επιτυχία. Αυτό μάλιστα το σημείο είναι και το πλέον σοβαρό μειονέκτημα των μεθοδολογιών βασισμένων σε χρήστες, ότι δηλαδή είναι από λίγο έως πάρα πολύ δύσκολο να βρει κανείς αντιπροσωπευτικό δείγμα χρηστών για αξιολόγηση σε πραγματικές συνθήκες εργασίας, πράγμα που με τη σειρά του συνεπάγεται αυξημένο κόστος αξιολόγησης και αυξημένη δυσκολία διεξαγωγής. Και αυτό ακριβώς το σημείο είναι που έχει κάνει τις μεθοδολογίες με ειδικούς τόσο δημοφιλείς, δηλαδή το ότι αίρουν αυτούς τους περιορισμούς, αντικαθιστώντας τους χρήστες με ειδικούς, με χαμηλότερα μεν ποιοτικά αποτελέσματα, αλλά και με σαφώς χαμηλότερο κόστος και πολύ χαμηλότερη δυσκολία διεξαγωγής της αξιολόγησης, άρα πολύ υψηλότερο δείκτη κόστους/απόδοσης. Συνεχίζοντας αυτό τον προβληματισμό οι Jeffries et al. (1991) συνοψίζουν τα σημαντικότερα μειονεκτήματα των δύο προσεγγίσεων ως εξής:

- Στις αξιολογήσεις με ειδικούς είναι δύσκολο να βρει κανείς τους σπάνιους ειδικούς του HCI, που είναι όμως απαραίτητοι
- Αν οι αξιολογητές μας δεν είναι μέλη της σχεδιαστικής ομάδας μπορεί να μην είναι ενήμεροι των τεχνολογικών περιορισμών (το θέμα αυτό θα συζητηθεί στο «Κεφάλαιο 2»),
- Μπορεί να εγερθούν και γενικότερες διαφωνίες μεταξύ των ειδικών (συναντήσαμε ένα τέτοιο πρόβλημα που συζητείται στο *Κεφάλαιο 2* στην Παράγραφο 2.9.)
- Στις εμπειρικές αξιολογήσεις πρέπει να έχουμε οπωσδήποτε ένα πρωτότυπο έτοιμο, με συνέπεια οι απόψεις των χρηστών πολλές φορές να έρχονται πολύ αργά, όταν πια δεν είναι δυνατό να αλλάξει τίποτα
- Τέλος οι εμπειρικές αξιολογήσεις καταναλώνουν πολλούς πόρους σε χρήμα και χρόνο.

Στις αξιολογήσεις με χρήστες οι Nielsen & Landauer (1993) υποστηρίζουν ότι μπορούμε να προβλέψουμε κατά προσέγγιση και τον αριθμό των προβλημάτων που θα ανευρεθούν, αφού αυτά είναι συνάρτηση του αριθμού των χρηστών, ως εξής:

$$U(i) = N (1 - (1 - \lambda)^i)$$

όπου: U(i) ο αριθμός των προβλημάτων ευχρηστίας που θα ανευρεθούν
 i ο αριθμός των χρηστών – αξιολογητών
 N ο συνολικός αριθμός προβλημάτων στη διασύνδεση
 λ η πιθανότητα ανεύρεσης κάθε ενός προβλήματος από κάθε έναν αξιολογητή

Οι τιμές των N και λ διαφέρουν ισχυρά από έργο σε έργο, όμως μπορούν να υπολογιστούν από δεδομένα και διαγράμματα που συσσωρεύονται όσο προχωράει το έργο κατασκευής του συγκεκριμένου συστήματος.

Συνοψίζοντας να δώσουμε συγκεντρωτικά τα υπέρ και κατά των αξιολογήσεων με ειδικούς και με χρήστες

	Αξιολόγηση Ειδικών (Expert-based)	Έλεγχος Ευχρηστίας (Usability testing)
Υπέρ	<ul style="list-style-type: none"> • Φτηνή μέθοδος • Γρήγορη • Εύκολη στην προετοιμασία και τη διεξαγωγή • Εφαρμόζεται σε όλα τα στάδια • Αποτιμούν τη σοβαρότητα των προβλημάτων 	<ul style="list-style-type: none"> • Αποκαλύπτει προβλήματα πραγματικών χρηστών • Αποδοτική, ακόμα και σε περίπλοκες διασυνδέσεις • Άμεση αποκάλυψη προβλημάτων
Κατά	<ul style="list-style-type: none"> • Δεν ανακαλύπτει όλα τα προβλήματα • Χρειάζεται έμπειρους αξιολογητές • Δυσκολία στην πρόταση λύσεων • Χάνει συχνά την εποπτεία του «όλου» 	<ul style="list-style-type: none"> • Ακριβή και δύσκολη στην υλοποίηση • Χρειάζεται αντιπροσωπευτικό δείγμα • Δυσκολία στην ανεύρεση χρηστών • Μπερδεύουν τα σοβαρά με τα δευτερεύοντα προβλήματα

Ένα από τα βασικά σημεία που εστιάζει αυτή η διατριβή, είναι και η βελτίωση της ποιότητας των αποτελεσμάτων των μεθοδολογιών με ειδικούς ώστε να μπορέσουμε να πλησιάσουμε και ποιοτικά τα αποτελέσματα των αξιολογήσεων με πραγματικούς χρήστες.

3. Αξιολογήσεις προσανατολισμένες στις διεργασίες (task-oriented) και ανεξάρτητες διεργασιών (task-free)

Ένας ακόμα ανεπίσημος διαχωρισμός των μεθοδολογιών που αναφέρονται εδώ είναι και αυτός των αξιολογήσεων προσανατολισμένων ή μη στις διεργασίες (tasks). Για παράδειγμα η γνωστική περιδιάβαση, οι αναλυτικές μέθοδοι και πολλές εμπειρικές μεθοδολογίες βασίζονται στις διεργασίες, ο αξιολογητής δηλαδή πρέπει πρώτα να βγάλει μια λίστα των διεργασιών προς αξιολόγηση και μετά να διεξάγει τη συνεδρία. Αντίθετα, η ευρετική αξιολόγηση και πολλές εμπειρικές μεθοδολογίες είναι ανεξάρτητες από τις διεργασίες και βλέπουν τη διασύνδεση σαν όλον. Υπάρχουν κάποια πλεονεκτήματα και μειονεκτήματα σε αυτές τις προσεγγίσεις. Σύμφωνα με τους Lewis & Rieman (1994) οι προσανατολισμένες σε διεργασίες μεθόδους έχουν σαν πλεονεκτήματα:

- Εστιάζουν σε προβλήματα της διασύνδεσης που θα συμβούν στη διάρκεια πραγματικής δουλειάς του χρήστη
- Δίνουν κάποια ιδέα για τη σοβαρότητα του προβλήματος σε σχέση με τη δουλειά που γίνεται
- Πολλά από τα προβλήματα που αποκαλύπτονται γίνονται μόνο ορατά μέσα στο πλαίσιο της διεργασίας, δηλαδή με την αλληλουχία των ενεργειών που πρέπει να γίνουν για την ολοκλήρωση της διεργασίας.

Όμως υπάρχουν και μειονεκτήματα:

- Καλυπτικότητα. Ποτέ δεν υπάρχει χρόνος για να αξιολογηθούν όλες οι διεργασίες που θα εκτελέσει ένας χρήστης μέσα στη διασύνδεση. Πάντα επιλέγεται ένα αντιπροσωπευτικό σετ (Karoulis, 2000).
- Δια-διεργασιακές διασυνδέσεις (cross-task interactions). Περισσότερο συχνά απ' ότι φαίνεται ο χρήστης διαπλέκει διάφορα σημεία στο interface. Κάθε διεργασία συνεπώς που

αξιολογείται μόνη της δεν αποκαλύπτει προβλήματα που εμφανίζονται στην αλληλεπίδρασή της με άλλες διεργασίες, πχ. προβλήματα συνέπειας, όπως διαφορετική ονομασία του ιδίου αντικειμένου σε διαφορετικά σημεία (για παράδειγμα «Σώσιμο» σε ένα σημείο και «Αποθήκευση» σε άλλο)

- Τείνουν να χάνουν την εικόνα του «όλου». Ασχολούμενοι με κάθε διεργασία ξεχωριστά, χάνεται η εποπτεία του όλου συστήματος και η «αίσθηση» της συνολικής λειτουργίας του.

Σαν συμπέρασμα η διατριβή αυτή προτείνει τη χρησιμοποίηση δύο μεθοδολογιών που να αλληλοκαλύπτονται. Το θέμα αυτό θα συζητηθεί στο *Κεφάλαιο 3*, όπου θα παρουσιαστεί η προσέγγιση της «Συνδυαστικής Αξιολόγησης».

4. Για τις αναλυτικές προσεγγίσεις

A) Η ανάλυση ενεργειών (action analysis) έχει δύο βασικές φάσεις (Lewis & Rieman, 1994):

- ◆ Στην πρώτη φάση πρέπει να αποφασιστεί ποιά φυσικά και νοητικά βήματα θα κάνει ο χρήστης για να ολοκληρώσει την υπό έρευνα διεργασία στη διασύνδεση
- ◆ Η δεύτερη φάση αναλύει αυτά τα βήματα ψάχνοντας για προβλήματα.

Τα προβλήματα που μπορεί να ανακαλύψει μια τέτοια αξιολόγηση είναι ότι:

- Χρειάζονται πολλά βήματα για την ολοκλήρωση μιας απλής διεργασίας
- Χρειάζεται πολύς χρόνος για να ολοκληρωθεί η διεργασία
- Χρειάζεται να μάθει ο χρήστης πολλά για τη διασύνδεση και άλλα τέτοιου είδους προβλήματα.
- «Οπές – holes» στη διασύνδεση, πράγματα δηλαδή που η διασύνδεση θα έπρεπε να μπορεί να κάνει, αλλά δεν κάνει.

B) Για τις αναλυτικές προσεγγίσεις αξιολόγησης, οι Lewis & Rieman (1994) διακρίνουν την επίσημη (keystroke-level analysis) και την ανεπίσημη μορφή (back-of-the-envelope).

Η *επίσημη μορφή* βασίζεται σε μετρήσεις που έχουν γίνει από ερευνητές σε εκατοντάδες χιλιάδες ενέργειες που γίνονται σε μια διασύνδεση (Olson & Olson, 1990). Για παράδειγμα, οι ερευνητές έχουν βρει ότι ο χρόνος που χρειάζεται για την πληκτρολόγηση ενός χαρακτήρα σε ένα στάνταρ πληκτρολόγιο είναι 0,28 sec, για να δείξει ο χρήστης με το ποντίκι ένα αντικείμενο στην οθόνη 1,5 sec, ενώ για να κινήσει το χέρι του από το πληκτρολόγιο στο ποντίκι 0,36 sec και για το χτύπημα ενός function key 0,21 sec. Βασιζόμενοι τώρα σε αυτές τις μετρήσεις μπορούμε να προβλέψουμε την επίδοση (σε χρόνο) του χρήστη κατά τη χρήση της συγκεκριμένης διασύνδεσης. Οι Lewis & Rieman (1992) υποστηρίζουν ότι αυτή η λεπτομέρεια στην αναλυτική μέθοδο δίνει αποτελέσματα τόσο ακριβή, ώστε να μπορούμε να προβλέψουμε το χρόνο εκτέλεσης μιας διεργασίας με περιθώριο σφάλματος 20%.

Το μειονέκτημα της επίσημης αναλυτικής μεθόδου είναι ότι είναι δύσκολη, ακριβή και απασχολεί σχετικά πολλούς πόρους. Για να καλυφθούν αυτά τα μειονεκτήματα οι Lewis & Rieman (1992, pp.41-42) προτείνουν την πρόχειρη (back-of-the-envelope) μέθοδο, μετά από παρατηρήσεις της δουλειάς του Kieras (1988) που προσπάθησε να απλοποιήσει το μοντέλο GOMS. Έχει και αυτή τις δύο φάσεις της καταλογοποίησης των ενεργειών και μετά της επισκόπησής τους, όμως εδώ δεν απαιτείται και πολύ αναλυτική λίστα ενεργειών, αλλά ενδιαφέρει η γενική εικόνα. Δημιουργείται μια «φυσική» σειρά ενεργειών και στη συνέχεια αξιολογείται. Οι ερευνητές αυτοί προτείνουν και μια σειρά ερωτήσεων, όπως:

- Μπορεί μια απλή διεργασία να γίνει με μια απλή σειρά ενεργειών;
- Μπορούν συχνά επαναλαμβανόμενες διεργασίες να γίνονται γρήγορα;

- Πόσες δράσεις και βήματα πρέπει να μάθει ο χρήστης;
- Περιγράφονται όλα στην τεκμηρίωση;

Αν θέλουμε να υπολογίσουμε και κάποιους χρόνους στην πρόχειρη προσέγγιση, τότε οι ερευνητές αυτοί υποστηρίζουν ότι θα υποθέτουμε ότι κάθε ενέργεια χρειάζεται δύο με τρία δευτερόλεπτα.

5. Το GOMS

Το Άγιο Δισκοπότηρο (Holy Grail) των επιστημόνων της ευχρηστίας είναι κατά τον Nielsen (1993b) η εφεύρεση αναλυτικών μεθόδων που θα επέτρεπαν στους σχεδιαστές να προβλέψουν την ευχρηστία της διασύνδεσης πριν ακόμα αξιολογηθεί. Προφανώς πρόκειται εδώ για αναλυτική μέθοδο όχι αξιολόγησης, αλλά του τομέα του σχεδιασμού ως προς ευχρηστία (usability engineering) και μάλιστα το ζητούμενο είναι μια θεωρία σχεδιασμού της διασύνδεσης βασισμένη στους στόχους του χρήστη. Η πιο πετυχημένη προσπάθεια μέχρι στιγμής είναι το μοντέλο GOMS – Goals, Operators, Methods, Selection rules των Card et al., (1983) και έρχεται από το χώρο της γνωστικής ψυχολογίας.

Το βασικό GOMS μοντέλο είναι απλό: Παρουσιάζει τους στόχους (*goals*) και υποστόχους του χρήστη (πχ. να αλλάξει μια λέξη σε ένα κείμενο), τους διαχειριστές (*operators*) που διαθέτουν οι χρήστες σαν κινητικές, αντιληπτικές ή γνωστικές οντότητες (πχ. κλικ με το ποντίκι, να κυττάξει στη μπάρα των μενού, να θυμηθεί ένα όνομα), τις μεθόδους (*methods*) που συνθέτουν οι χρήστες με αλληλουχίες από αυτούς τους διαχειριστές για να πετύχουν το στόχο τους, (πχ. επιλογή της λέξης με κίνηση του ποντικιού πάνω της και διπλό κλικ), και τους κανόνες επιλογής (*selection rules*) που είναι απαραίτητοι όταν υπάρχουν περισσότερες επιλογές ή διαδοχικές ενέργειες για την ολοκλήρωση του στόχου (πχ. η λέξη μπορεί να απομακρυνθεί με την επιλογή της και στη συνέχεια με την εντολή «Αποκοπή – Cut» ή με το πλήκτρο *backspace*). Θεωρητικά έτσι θα είναι δυνατόν για κάθε ενέργεια να υπολογιστούν χρόνοι και επιδόσεις χρηστών. Για παράδειγμα ο νοητός διαχειριστής (*mental operator*) του να θυμηθεί κανείς ένα όνομα αποτιμάται από ψυχολογικά πειράματα στα 1,35 δευτερόλεπτα.

Προφανώς όμως το να μοντελοποιήσει κανείς έτσι μια διασύνδεση θα ήταν και πολύ δύσκολο, αλλά και πολύ μεγάλο σε έκταση. Άλλη αδυναμία του μοντέλου είναι (Carroll & Campbell, 1986) ο περιορισμός ότι οι χρήστες είναι αλάνθαστοι. Στην πράξη όμως και οι αρχάριοι και οι περιστασιακοί, αλλά και οι έμπειροι χρήστες κάνουν πολλά λάθη, θέμα που είναι πρωταρχικής σημασίας στο σχεδιασμό ως προς ευχρηστία. Έχουν γίνει τροποποιήσεις στο μοντέλο (Olson & Olson, 1990) με σκοπό την απάλειψη αυτών των αδυναμιών, ενώ επίσης έρευνες, όπως πχ. των Gray et al. (1992) και Nielsen & Phillips (1993), έχουν δείξει επιτυχή εφαρμογή του μοντέλου στον πραγματικό κόσμο.

Το θέμα τελικά όμως είναι ότι ούτε και το GOMS δεν μπορεί να δώσει τελικές λύσεις στο σχεδιασμό ως προς ευχρηστία, όμως, λόγω της ισχυρής του θεωρητικής θεμελίωσης, έδωσε λαβή για άλλες προσεγγίσεις, όπως η προσέγγιση του στόχου – διεργασίας – ενέργειας (*goal – task – action*) που ακολουθείται από πολλές μεθοδολογίες αξιολόγησης και περιγράφεται στην παρούσα διατριβή στο «Κεφάλαιο 2: Η Παράμετρος του Χρόνου στην Αξιολόγηση – CGJ».

Κλείνοντας αυτό το κεφάλαιο αφήνουμε ίσως αναπάντητα πολλά ερωτήματα, τόσο για την καταλληλότερη μέθοδο κάθε φορά, όσο και για την αξιοπιστία της μεθόδου που τελικά θα επιλέξουμε. Σαφής απάντηση δεν υπάρχει σε καμμία από αυτές τις ερωτήσεις, όμως υπάρχει ο «σωστός δρόμος» που δεν είναι άλλος από τη σωστή προετοιμασία της αξιολόγησης: αν ακολουθηθούν σωστά οι οδηγίες προετοιμασίας που αναφέρθηκαν σ' αυτό το κεφάλαιο, η αξιολόγησή μας έχει πολλές πιθανότητες να είναι επιτυχημένη.

6. Βιβλιογραφικές Αναφορές (References)

- (Αβούρης, 2000) Αβούρης, Ν. *Εισαγωγή στην Επικοινωνία Ανθρώπου-Υπολογιστή*. Εκδ. Διάυλος. Αθήνα, 2000
- (Aedo et al., 1996) Aedo, I., Catenazzi, N. & Diaz, P. The Evaluation of a Hypermedia Learning Environment: The CESAR Experience. *Journal of Educational Multimedia & Hypermedia*, 1996, 5(1), pp. 49-72
- (Benyon et al, 1990) Benyon, D., Davies, G., Keller, L., and Rogers, Y. *A Guide to Usability-Usability now!* Milton Keynes: The Open University, 1990
- (Berry, 1988) Berry, R.E. Common User Access – A Consistent and Usable Human-Computer Interface for the SAA Environment. *IBM Systems Journal*, 27, 3, pp281-300
- (Bias, 1991) Bias, R Walkthroughs: Efficient Collaborative Testing. *IEEE Software*, 8, 5 (September), pp. 94-95
- (Bransford et al., 1990) Bransford, J.D., Sherwood, R.D, Hasselbring, T.S. Kinzer, C.K and Williams, S.M. Anchored Instruction: Why we Need it and How Technology Can Help. In Nix, D.Spiro, R. (Edts.) *Cognition, Education, Multimedia*. Lawrence Erlbaum Ass. Hillsdale, 1990. pp. 115-141
- (Brooke et al., 1990) Brooke, J., Bevan, N., Brigham, F.R. Harker, S. and Youmans, D. Usability Assurance and Standarization – Work in Progress in ISO. *Proc. IFIP INTERACT '90 Third Intl. Conf. On Human-Computer Interaction*, Cambridge, UK, 27-31 August 1990, pp 357-361
- (Calder, 1995) Calder, J. *Programme evaluation and quality: a comprehensive guide to setting up an evaluation system*, Institute of Educational Technology, Open University, Kogan Page, London, 1995
- (Caplan, 1990) Caplan, S. Using Focus Groups Methodology for Ergonomic Design. *Ergonomics*, 33, 5, pp. 527-533
- (Card et al, 1983) Card, S.K., Moran, T.P. and Newell, A. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1983
- (Carroll & Campbell, 1986) Carroll, J.M., and Campbell, R.L. Softening up Hard Science: Reply to Newell and Card. *Human-Computer Interaction*, 2, 3, pp.227-249
- (Catenazzi et al., 1997) Catenazzi, N., Aedo, I., Diaz, P. & Sommaruga, L. The Evaluation of Electronic Book Guidelines from two Practical Experiences. *Journal of Educational Multimedia & Hypermedia*, 1997, 6 (1), pp. 91-114
- (Chinapah & Miron, 1990) Chinapah, V., and Miron, G. *Evaluating educational programmes and projects: holistic and practical considerations*, Socio-economics Studies, No 15, UNESCO, Paris, 1990.
- (Δημητριάδης, 2000) Δημητριάδης, Στ. *Τεχνολογία των Πολυμέσων και η Εφαρμογή της στην Εκπαιδευτική Διαδικασία*. Διδακτορική Διατριβή, ΑΠΘ, τμήμα Πληροφορικής, 2000.
- (Δημητρόπουλος, 1991) Δημητρόπουλος, Ε. *Εκπαιδευτική αξιολόγηση: η αξιολόγηση της εκπαίδευσης και του εκπαιδευτικού έργου*, εκδ. Γρηγόρη, Αθήνα, 1991.
- (Diaper, 1989) Diaper, D. Task Observation for Human-Computer Interaction. In Diaper, D. (Edt.) *Task Analysis for Human-Computer Interaction*. Ellis Horwood, Chichester, UK, 1989, pp. 210-237
- (Duffy & Knuth, 1990) Duffy, T.M. and Knuth, R.A. Hypermedia nad Instruction. Where is the Match?. In Jonassen, D.H. and Mandl, H. (Edts.) *Designing Hypermedia for Learning*. NATO ASI Series F, vol. 67, Springer Verlag, Germany, 1990.

- (Ericsson & Simon, 1984) Ericsson, K.A. and Simon, H.A. *Protocol Analysis: Verbal Reports as Data*. The MIT Press, Cambridge, MA, 1984
- (Georgiadou & Economides, 2000) Georgiadou, E., and Economides, A. Evaluation Factors of Educational Software. *Proc. of IWALT 2000 conf.* 4-6 Dec 2000 Palmerston North, New Zealand, pp. 113-116
- (Goldman & McDonald, 1987) Goldman, A.E., and MacDonald, S.S. *The Group depth Interview: Principles and Practice*. Prentice-Hall, Englewood Cliffs, NJ, 1987
- (Good, 1989) Good, M. Developing the XUI Style. In Nielsen, J. (Edt.) *Coordinating User Interfaces for Consistency*, Academic Press, Boston, MA, pp.57-73
- (Gray et al., 1992) Gray, W.D., John, B.E. and Atwood, M.E. The Precip of Project Ernestine, or, an Overview of a Validation of GOMS. *Proc. of ACM CHI '92 conf.*, Monterey, CA, 3-7 May, 1992, pp.307-312
- (Greenbaum, 1988) Greenbaum, T.L. *The Handbook for Focus Group Research*. D.C. Heath & Co., Lexington, MA, 1988
- (Greenbaum, 1993) Greenbaum, T.L. *The Practical Handbook and Guide to Focus Group Research*. Lexington Books, New York, NY, 1993
- (Hewett & Scott, 1987) Hewett, T.T. and Scott, S. The Use of Thinking-Out-Loud and Protocol Analysis in Development of a Process Model of Interactive Database Searching. *Proc. of IFIP INTERACT '87 Second Intl. Conf on Human-Computer Interaction*, Stuttgart, Germany, 1-4 September, 1987, pp. 51-56
- (ISO, 1998) ISO 9241 - International Standardization Organization. *Ergonomic requirements for office work with visual display terminals (VDT's)*, 1998.
- (Jeffries et al., 1991) Jeffries, R.J., Miller, J.R., Wharton, C., and Uyeda, K.M. User Interface Evaluation in the Real World: A Comparison of four Techniques. *Proc. of CHI '91*, ACM, New York, 1991, pp.119-124
- (Kantner & Rosenbaum, 1997) Kantner, L, and Rosenbaum, S. Usability Studies of WWW Sites: Heuristic Evaluation vs. Laboratory Testing. *Proc of SIGDOC 97 conf.*, Snowbird, Utah, USA, 1997
- (Karat et al., 1992) Karat, C., Campbell, R. & Fiegel, T. Comparison of Empirical Testing and Walkthrough Methods in User Interface Evaluation, 1992. *Proceedings of ACM CHI '92*. Monterey, CA, May 3-7, pp.397-404.
- (Karoulis et al, 2000) Karoulis, A., Demetriades, S., Pombortsis, A. The Cognitive Graphical Jogthrough – An Evaluation Method with Assessment Capabilities. *Applied Informatics 2000 Conference Proceedings*, February 2000, Innsbruck, Austria
- (Karoulis et al, 2000b) Karoulis, A., Demetriades, S., and Pombortsis, A. Evaluation of multimedia educational interfaces for the junior highschool classes using two different methodologies: the “Perivallon” experience. *Journal of Educational Multimedia and Hypermedia* (Under Submission)
- (Κασσωτάκης, 1990) Κασσωτάκης, Μ. *Η αξιολόγηση της επίδοσης των μαθητών*. εκδ. Γρηγόρη, Αθήνα, 1990.
- (Kato, 1986) Kato, T. What «Question-Asking Protocols» can say about the User Interface. *Intl. Journal of Man-Machine Studies*, 25, 6 (December), 1986, pp. 659-673
- (Kennedy, 1989) Kennedy, S. Using Video in the BNR Usability Lab. *ACM SIGCHI Bulletin*, 21, 2 (October), 1989, pp. 92-95

- (Kieras, 1988) Kieras, D.E. Towards a Practical GOMS Model Methodology for User Interface Design. In M. Helander (Ed.), *Handbook of Human-Computer Interaction*. Amsterdam: Elsevier Science (North-Holland), 1988
- (Lewis, 1982) Lewis, C. Using the Thinking-Aloud Method in Cognitive Interface Design. *IBM Research Report RC 9265*, Yorktown Heights, NY, 1982
- (Lewis et al., 1990) Lewis, C., Polson, P., Wharton, C. & Rieman, J. Testing a Walkthrough methodology for Theory-Based Design of Walk-Up-and-Use Interfaces. *Proceedings of ACM CHI '90, Seattle, Washington*. April 1-5, 1990, pp. 235-242
- (Lewis & Rieman, 1994) Lewis, C. and Rieman, J. *Task-centered User Interface Design - A practical introduction*, 1994. ftp.cs.colorado.edu/pub/cs/distribs/HCI-Design-Book
- (Mack & Burdett, 1992) Mack, R.L., and Burdett, J.M. When Novices Elicit Knowledge: Question-Asking in Designing, Evaluating and Learning to Use Software. In Hoffman, R. (Edt.) *The Psychology of Expertise: Cognitive Research and Empirical AI*. Springer Verlag, New York, NY, 1992, pp. 245-268
- (Μακροάκης, 1999) Μακροάκης, Β. Αξιολόγηση Συστημάτων Ανοικτής και εξ' Αποστάσεως Εκπαίδευσης. Στο *Ανοικτή και εξ' Αποστάσεως Εκπαίδευση, Τόμος Α'*, Ελληνικό Ανοικτό Πανεπιστήμιο, Πάτρα, 1999. pp.245-301
- (Maulsby et al., 1993) Maulsby, D., Greenberg, S., and Mander, R. Prototyping an Intelligent Agent Through Wizard of Oz. *Proc. of ACM InterCHI '93 conf*, Amsterdam, The Netherlands, 24-29 April, 1993. pp. 277-284
- (Muller et al., 1998) Muller, M.J., Matheson, L., Page, C., and Gallup, R. Participatory Heuristic Evaluation. *Interactions*, vol.5 (Sep-Oct. 98), 1998
- (Nielsen, 1990) Nielsen, J. Big Paybacks from «discount» usability engineering. *IEEE Software*. 7, 3 (May), 1990. pp.107-108.
- (Nielsen, 1993b) Nielsen, Jacob. *Usability Engineering*. Academic Press, San Diego, 1993.
- (Nielsen, 2001a) Nielsen, J. www.useit.com May 2001
- (Nielsen & Landauer, 1993) Nielsen, J., and Landauer, T.K. A Mathematical Model of the Finding of Usability Problems. *Proc. of ACM INTERCHI '93 conf*, Amsterdam, The Netherlands, 24-29 April, 1993, pp. 206-213
- (Nielsen & Mack, 1994) Nielsen, J. and Mack, R.L. (eds.) *Usability Inspection Methods*. John Wiley & Sons, 1994
- (Nielsen & Molich, 1990) Nielsen, J., and Molich, R. Heuristic Evaluation of User Interfaces, *Proc. of Computer-Human Interaction Conference (CHI)*, Seattle, WA, 1-5 April, 1990, p. 249-256
- (Nielsen & Phillips, 1993) Nielsen, J., and Phillips, V.L. Estimating the Relative Usability of Two Interfaces: Heuristic, Formal, and Empirical Methods Compared. *Proc. of ACM INTERCHI '93 conf*. Amsterdam, The Netherlands, 24-29 April 1993, pp.214-221
- (O'Donell et al., 1991) O'Donell, P.J., Scobie, G., and Baxter, I. The use of Focus Groups as an Evaluation Technique in HCI. In Diaper, D. and Hammond, N. (Edts.) *People and Computers VI*, Cambridge University Press, Cambridge, UK, pp.211-224
- (O'Malley et al., 1984) O'Malley, C.E., Draper, S.W., and Riley, M.S. Constructive Interaction: A Method for Studying Human-Computer-Human Interaction. *Proc. of IFIP INTERACT '84 First Intl. Conf. on Human-Computer Interaction*, London, UK, 4-7 September, 1984, pp. 269-274.

- (Olson & Olson, 1990) Olson, J.R, and Olson, G.M. The Growth of Cognitive Modeling in Human-Computer Interaction since GOMS. *Human-Computer Interaction*, 5, 1990, pp. 221-265
- (OU, 2000) *Evaluation Methods and Procedures for Studying Learners' Use of Media*, Institute of Educational Technology, The Open University, Milton Keynes, April 2000, <http://iet.open.ac.uk/PLUM/evaluation/contents.html>
- (Preece et al, 1994) Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S., Carey, T., *Human-Computer Interaction*, Addison-Wesley Publishing Company, 1994.
- (Polson et al, 1992) Polson, P.G., Lewis, C., Rieman, J. and Warton, C. Cognitive Walkthroughs: a Method for Theory-based Evaluation of User Interfaces. *International Journal of Man-Machine Studies*, 36, pp.741-773, 1992
- (Race, 1993) Race, Phil. *The Open Learning Handbook*. Kogan Page, London, 1993. Ελληνική έκδοση *Το Εγχειρίδιο της Ανοικτής Εκπαίδευσης*. Μεταίχμιο, 1999
- (Roth & Chair, 1997) Roth, W., & Chair, L. Phenomenology, Cognition and the design of Interactive Learning Environments. *Proceedings of ED-MEDIA & ED-TELECOM 97, Calgary, Canada, 1997*. Association for the advancement of Computing in Education (AACE), Charlottesville, VA., pp. 1101-1107
- (Rowley & Rhoades, 1992) Rowley, D. & Rhoades, D. The Cognitive Jogthrough: A Fast-Paced User Interface Evaluation Procedure, 1992. *Proceedings of ACM CHI '92, Monterey, California, May 3-7, 1992*, pp. 389-395
- (Scriven, 1976) Scriven, M. "The methodology of evaluation". In *Perspectives of Curriculum Evaluation*, Tyler, R. (edt.), Rand McNally, Chicago, 1967
- (Shneiderman, 1998) Shneiderman, Ben, *Designing the User Interface, 3rd ed.*, Addison-Wesley Publishing Company, 1998
- (Spiro & Jehng, 1990) Spiro, R.J. and Jehng, J. Cognitive Flexibility and Hypertext: Theory and Technology for the Nonlinear and Multidimensional Traversal of Complex Subject Matter. In Nix, D. and Spiro, R. (Edts.) *Cognition, Education, Multimedia*. Hillsdale, Erlbaum, New Jersey, 1990.
- (Stewart, 1990) Stewart, T. SIOIS – Standard Interfaces or Interfaces Standards. *Proc. IFIP INTERACT '90 Third Intl. Conf. On Human-Computer Interaction*, Cambridge, UK, 27-31 August 1990, pp xxix-xxxiv
- (Tognazzini, 1989) Tognazzini, B. Achieving Consistency for the Macintosh. In Nielsen, J. (Edt.) *Coordinating User Interfaces for Consistency*, Academic Press, Boston, MA, pp.57-73.
- (Vosniadou, 1996) Vosniadou, S. Learning Environments for Representational Growth and Cognitive Flexibility. In Vosniadoy, S, de Corte, E., Glaser, R., and Mandl, H. (Edts.) *International Perspectives on the Design of Technology-Supported Learning Environments*. Lawrence Erlbaum Ass., Mahwah, New Jersey, 1996. pp. 13-24
- (Wharton et al., 1992) Wharton, C., Bradford, J., Jeffries, R. & Franzke, M. Applying Cognitive Walkthroughs to More Complex User Interfaces: Experiences, Issues and Recommendations, 1992. *Proceedings of ACM CHI '92. Monterey, CA, May 3-7*, pp. 381-388
- (Wharton et al., 1994) Wharton, C., Rieman, J., Lewis, C. and Polson, P. The Cognitive Walkthrough: A Practitioner's Guide. In Nielsen, J. and Mack, R.L. (edts) *Usability Inspection Methods*. New York: John Wiley & Sons, 1994
- (Wixon et al., 1990) Wixon, D., Holtzblatt, K. and Knox, S. Contextual Design: an emergent view of system design. In *Empowering people*, CHI '90 conference proceedings (Chew, J.C., and Whitwside, J., eds) pp. 329-336. New York, ACM Press, 1990.

(Wolf, 1989) Wolf, R. Consistency As Process. In Nielsen, J. (Edt.) *Coordinating User Interfaces for Consistency*, Academic Press, Boston, MA, pp.89-92