Phonemic-graphemic correspondence hypotheses

The following hypotheses are based on the following assumptions:
- that we all store all our words in our minds; each person has their own personal store, which might be called their 'mental lexicon'
- that we all store for each word, among other things, a pronunciation and a spelling
- that we are all aware of the pronunciation and spelling of each word in our mental lexicon and can call them to mind
- that most linguistically untrained people are not aware of the phonological and orthographical systems that operate in their minds that regulate the pronunciation and spelling of each word
- that there is a small minority of words in each person's lexicon which each person feels uncertain about in either pronunciation or spelling, or both; but these uncertainties do not undermine the existence and operation of either the phonological or the orthographical system
- that the orthographical system derives from the phonological system in all cultures that use a non-semantic writing system, ie those that use alphabets, syllabaries or some form of rebus writing
- that just as it is possible to investigate the nature and operation of a phonological system, it is also possible to investigate the nature and operation of a corresponding orthographical system

The following hypotheses relate to the English orthographic system for British educated young adults. They have been set out as a basis for investigating the nature and operation of orthographic competence as a resource for spelling novel words which conform to the phonological structure of British English words, ie the productive orthographic system in the mind. The hypotheses are based on high frequency, regular, conventional correspondences between phonemes, phonemic sequences and syllable structure on the one hand and orthographic correspondences on the other; we believe that, as a general rule, this productive orthographic system will ignore low frequency, irregular, idiosyncratic spellings, such as
- for /k/ and /kw/ *liquor, liqueur, lacquer, choir*
- 'silent' letters as in *wrap, white, two, who*
- etc

but it will not ignore a perception of foreign loan words and technical terms, for which different systems exist.

It should be noted that the hypotheses do not correspond to a phonemic-style re-spelling transcription system as exists, for instance, in the *Oxford BBC Guide to Pronunciation* (OUP, 2006). Nor do they correspond to the type of 'regularized spelling' that is advocated for language teaching purposes – see, for example, Derwing, Priestly & Rochet 'The Description of Spelling-to-Sound Relationships in English, French and Russian', in Luelsdorff, P A (1987) *Orthography and Phonology* Amsterdam: John Benjamins.

Rather, the hypotheses seek to represent the **Sound-to-Spelling** correspondences that exist as a system in the mind, which enables a native speaker/writer to **do** something,

namely to spell novel, native-like words. The phonemic inventory that they are based on is that of Southern England Standard Pronunciation (formerly known as 'RP'):

consonants

plosives:	p, b, t, d, k, g
nasals:	m, n, ŋ
fricatives:	f, v, θ, ð, s, z, ʃ, ʒ, h
affricates:	tʃ, dʒ
approximants:	l, r, j, w

vowels: strong

short:	ɪ, ɛ, æ, ɒ, ʊ, ʌ
long monophthongs:	iː, ɑː, ɔː, uː, ɜː
long diphthongs:	eɪ, aɪ, ɔɪ, əʊ, aʊ, ɪə, ɛə, ʊə
weak:	i, u, ə

stress
primary and secondary stress, and no ('weak') stress

The orthographic system is presented as sets of 'graphemes':

'protographemes' (single letters):	a, b, c, d, e, etc
which may be 'doubled':	bb, cc, dd, ee, etc
'bigraphemes' (regular pairs for single phonemes):	ch, sh, th, ai, ay, ea, er, ge, etc
'trigraphemes' (regular trios for single phonemes):	air, are, ear, eer, oor, our, etc
'quadrigrapheme' (regular quad for single phoneme):	aire

Thus a 'grapheme' may consist of one or more letters to represent a single sound. One 'quasi-protographeme' (single letter that corresponds to a phoneme sequence) is recognized: <x> for /ks/ or /gz/; and one 'quasi-bigrapheme' (regular pair for a phoneme sequence) is recognized: <qu> for /kw/.

Phonemes are represented as: /b/; graphemes as <b>. The above terminology derives from Sgall, P 'Towards a Theory of Phonemic Orhtography' in Luelsdorff, P A (1987) *Orthography and Phonology* Amsterdam: John Benjamins.

The hypotheses are presented as
- predictions ("that they will…"; 67% - 100%)
- options ("that they will … either …or" …; 33% - 66%)
- possibilities ("that they may …"; 1% - 32%)

General hypotheses

G1   that adult native speakers of English will utilize all the letters of the Roman alphabet as used for Standard English for attempting to spell novel words that conform to English phonological systems

G2   that they will use the consonant letters <b, c, d, f, g, h, j, k, l, m, n, p, r, s, t, v, w, x, y, z> as single graphemes ('protographemes')

G2a   that they will also use the consonant letters <b, c, d, f, g, l, m, n, p, r, s, t, z> doubled in certain contexts ('double graphemes')

G2b   that they will not use the consonant letters <h, j, k, v, w, x, y> doubled

G2c   that they will use the following consonant letter pairs ('bigraphemes') to represent single phonemes: <ch, ph, sh, th> and <ck, dg>, and may use <wh>

G2d   that they will use <tch> as a 'trigrapheme'

G3   that they will use the vowel letters <a, e, i, o, u> as single graphemes

G3a   that they will use the vowel letters <e, o> doubled to represent single phonemes

G3b   that they will not use the vowel letters <a, i, u> doubled

G3c   that they will use the following vowel letter pairs ('bigraphemes') to represent single phonemes: <ai, au, ea, ia, ie, oa, oe, ou, ua, ue>

G4   that they will also use combinations of vowels and consonants to represent single phonemes

G4a   that they will use each vowel letter with <r> as pairs ('bigraphemes') to represent single phonemes: <ar, er, ir, or, ur>

G4b   that they will use the following pairs ('bigraphemes') in word-final position: <aw, ay, ew, ow, oy; ge> and also the following sequences ('trigraphemes' and one 'quadrigrapheme') in word-final position: <dge, air, are, ear, eer, ere, ire, oar, oor, ore, our, ure; aire>

G5   that they will use <x> as a 'quasi-protographeme' to represent /ks, gz/ and <qu> as a 'quasi-bigrapheme' to represent /kw/

G6   that they will use a minimum of three letters for any word that is perceived to be a lexical item, even though grammatical items may consist of only one or two letters – the 'three letter rule'. (Compare *I/eye*; *a/say*; *at/add*; *to/toe*; *be/bee*, etc; but note *do, go* as lexical as well as grammatical items.)

G6a   that they may use a minimum of two letters for a word perceived to be a foreign loan word or a technical term: *id, pi*

G7   that they will not indicate stress

G8   that they will distinguish a spelling system for words that are perceived to conform to common English words from those that are perceived to conform to loan words from another language or technical words.

G9   that they will operate a spelling system for parts of words that conform to perceived morphemes

G10   that they may use a hyphen <-> as a link between a morpheme with a perceived final vowel letter and a morpheme with a perceived initial vowel letter, eg
co-operate, de-ice, psycho-analysis

Specific hypotheses

Consonants
Evidence will be taken as far as possible from morphologically simple words. Graphemes are hypothesized according to phonological environments: word-initial; medial: following either a long, or a short, or a weak vowel, and preceding either a vowel or a consonant; word-final: following either a long, or short, or a weak vowel, or consonant.

C1   that for /**p, b, t, d**/ they will use <p, b, t, d> in word-initial position, after a long vowel, after a perceived spelling of a vowel bigrapheme, after a weak vowel, before a consonant, and in word-final position, except in the case of a monosyllabic lexical item that has no onset consonant.
pat, paper, staple, couple, callipers, multiple, tap, apt, tulip;
bat, stable, obtain, cannibal, vegetable, tab, cherub
tap, hotel, title, threat, rickety, marital, pat, waltz, ticket;
dare, tidy, idle, adjunct, parody, red, stupid

C1a   that they will use double <p, b, t, d> in medial position after a short vowel, and in word-final position of monosyllabic lexical item that has no onset consonant.
happy; *abbey*; *otter*; *adder*; *ebb*; *add* ('three letter rule')

C1b   that they may use double <p, b, t, d> in medial position after a word-initial weak vowel
apparent, oppose, supply (but NB *repeat, surprise*); *abbreviate*
attack, address

C2   that for /**k**/ they will use <k> before <e, i, y> and <c> before other vowels and consonants in initial position
kettle, kill, shaky; *cattle, cull, vocal; act*

C2a   that they will use this distribution for /k/ after a long vowel (and after a perceived spelling of a vowel bigrapheme) and a weak vowel in medial position
beaker; *bacon, focus; treacle; treadle*
miracle, obstacle, monocle, vehicle
electrical, critical, musical, physical

C2b   that they will use <ck> after a short vowel in medial position
beckon, trickle

C2c   that they will use either <c> or <k> after a short vowel and <n>, before <-le>
uncle, wrinkle, ankle

C2d   that they will use <k> after a long vowel (and after a perceived spelling of a vowel bigrapheme) and <ck> after a short vowel in word-final position
break, back; book

C2e   that they will use <ck> after the weak vowel /ə/ in final position, but <c> after the weak vowel /ɪ/
haddock, hammock, (but NB *havoc*)

electric, critic, music, mimic (but NB *gimmick*)

C2f   that they may use double <c> for /k/ in medial position after a word-initial weak vowel
account, accord, accommodate

C2g   that for /kw/ they will use <qu>
quick, frequent

C2h   that for /ks/ they will use <x> in morphologically simple words, but <cs> in a final unstressed syllable in a word denoting study
extra, excellent, axle, flaxen, hoax, box (NB *ox* with two letters only, but *axe* with 'three letter rule' in British English)
physics, electronics

C3   that for /**g**/ they will use <g> before all vowels and consonants, although Consonant Hypothesis 15 applies in some cases before <e, i, y>
gas, get, give, go, gum, great, glean (NB but *gem, gin, gym*)

C3a   that they may use <g> after long vowels in medial position
eager, tiger, eagle, ogre (NB but *danger*)

C3b   that they will use <gg> after short vowels in medial position, and in word-final position of monosyllabic lexical item that has no onset consonant.
baggage, wriggle, egg

C3c   that they will <gue> after long vowels in word-final position
vague, rogue

C3d   that they will use <g> after short vowels in word-final position except in the case of a monosyllabic lexical item that has no onset consonant.
bag, jig

C3e   that they will use <x> for /gz/ before a vowel
exist, exhibit, inexorable

C4   that for /**m, n**/ they will use <m, n> in word-initial position, after a long vowel, after a perceived spelling of a vowel bigrapheme, before a consonant, and in word-final position, except in the case of a monosyllabic lexical item that has no onset consonant.
mat, rumour, rum, rhythm; *nut, lunar, fiend, find, run, happen*

C4a   that they will use double <m, n> in medial position after a short vowel, and in word-final position of monosyllabic lexical item that has no onset consonant.
summer, pummel; *funny, funnel, inn* ('three letter rule')

C4b   that they may use double <m, n> in medial position after a word-initial weak vowel
immense, innate (but NB *imagine, inert*)

C5 that for /ŋ/ before a consonant, they will use <n>, but in final position, they will use <ng>
anchor, anxious, anger, angle, hang, hanging

C6 that for /**f**/ they will use <f> in word-initial position, after a long vowel, after a perceived spelling of a vowel bigrapheme, before a consonant, and in word-final position of grammatical items
fan, wafer, rifle, leaf, deaf, lift, if

C6a that they will use double <f> in medial and final positions after a short vowel
suffer, waffle, stiff, off

C6b that they may use double <f> in medial position after a word-initial weak vowel
affair, affirm (but NB *refuse*)

C7 that for /**v**/ they will use <v> always, will never double it, and will never use it in word-final position
van, over, oven, stove, serve, have

C8 that for both /**θ, ð**/ they will use <th> always in initial and medial positions
thin, than; ether, method, father, mother; anthem, brethren

C8a that they may use <the> to distinguish /ð/ from /θ/ in final position
bath/bathe; teeth/teethe (but NB *smooth, with; truth/truths; path/paths*)

C9 that for /**s**/ they will use <s> in word-initial position, and after a long vowel in medial position, but may use <c> before <e, i, y>
sand, sell/cell; siege/ceiling, song, sun, syphon/cycle; mason, basin, acid, recent

C9a that they may use double <s> or <sc> in medial position after a word-initial weak vowel
assent/ascent; associate (but NB *research*(vb); *receive*)

C9b that they will use <s> before a consonant
last, ask, latest

C9c that they will use double <s> in medial position after a short vowel, and in final position after a short vowel and the long vowel perceived as <a> in Southern England Standard Pronunciation
lesson, hassle; loss, lass, glass; ass

C9d that they may use <se> or <ce> for word-final /s/ after a long vowel or consonant
loose, lace; sparse, scarce; dense, pence

C9e that they may use <ss> or <ce> for word-final /s/ after a weak vowel, but <s> after a perceived spelling of a vowel bigrapheme
actress, useless, harness, happiness; malice, palace; callous

C10 that for /**z**/ they will use <z> in initial position
zoo, zone

C10a that they will use either <s> or <z> for /z/ in medial position, except after a short vowel before syllabic /l/
laser, positive, position, present(n, v or adj)*, weasel, measles, reason, risen; razor, hazard, wizard, dozen, frozen*

C10b that they will use double <z> after short vowels before syllabic /l/
dazzle, drizzle, nozzle, muzzle

C10c that they will use either <se> or <ze> in final position after long vowels
prise, prize, cause, gauze, ease, raise, rise, rose, ruse, size

C10d that they will use double <z> in final position after short vowels
jazz, fizz (but NB *quiz*)

C10e that they will use <s> in final /-ɪzm/
prism, racism

C11 that for /ʃ/ they will use <sh> in initial and final positions
shun, shame, wash, leash

C11a that they will use <t> after long vowels and consonants, and <ss> after short vowels for /-ʃən/ as a noun for a process or object (orthographic *–tion*, *-ssion*)
nation, completion, notion, solution, portion, caution; auction, action, suction, sanction
passion, mission, session, concussion

C11b that they will use <c> after short vowels for /-ʃən/ as a noun for a profession (orthographic *–cian*)
musician, physician

C11c they will use either <t> or <c> after vowels and <t> after consonants for /-ʃəl/ as an adjective (orthographic *–tial, -cial*)
initial, official, potential

C11d that they will use <ss> after short vowels for /-ʃə/ as a noun for a process or object (orthographic *–ure*)
fissure, pressure

C11e that they will use <sh> otherwise in medial position
fashion, threshold, bishop, worship, usher, cushion

C12 that for /ʒ/ they will use <s> in medial position
vision, fusion, treasure, closure, usual (but NB *seizure*)

C12a that they will use <ge> for /ʒ/ in final position

rouge, beige, collage, prestige

C13   that for /**h**/ they will use <h>
have, behave

C14   that for /**tʃ**/ they will use <ch> in initial position
chin, chain

C14a   that they will use either <t> before <u>, for /tʃə, tʃu-/, or <ch> after long vowels (and after a perceived spelling of a vowel bigrapheme) and consonants, and <tch> after short vowels
nature, natural, fracture, factual
orchard, treachery, pilchard; satchel, hatchet, wretched , pitcher, butcher

C14b   that they will use <ch> for /tʃ/ after long vowels and consonants in final position, and in grammatical items, and <tch> after short vowels in lexical items
beach, search, bench, belch; each, such, much, which
batch, fetch, witch, notch, hutch

C15   that for /**ðʒ**/ they will use <j> before all vowels, but may use <g> before <e, i, y> in initial position
jam, jet, jig, jog, jug; gem, gin, gym

C15a   that they will use <j>, or <g> before <e, i, y> after long vowels, weak vowels and consonants in medial position and <dg> after short vowels
major, majesty; wager, danger, surgeon, sergeant, margin, region, orgy, strategy, energy, vengeance
badger, ledger, fidget, budget (but NB *rigid*)

C15b   that they will use <ge> after long vowels, weak vowels and consonants in final position, and <dge> after short vowels
age, change, bilge; badge, hedge, ridge, lodge, judge; village, privilege

C16   that for /**l**/ they will use <l> in initial position, before a consonant and in medial and final positions after a long vowel (and after a perceived spelling of a vowel bigrapheme) and after a weak vowel in medial position
long; health; silent; fool; wool; family

C16a   that they will use double <l> in medial and final positions after a short vowel
silly; fill, full

C16b   that they will use <le> for syllabic /l/ or /əl/ in final position, or <al> for most adjectives
idle, able, little, bottle, mettle, peddle, muscle (but NB *model, morsel, idol*)
vital, modal, electrical, metal, pedal

C17   that for /**r**/ they will use <r> in initial position, in medial position after a long vowel or a consonant, and after a weak vowel
red; wary, dairy; walrus; salary, memory, diary

C17a   that they may use either <r> or double <r> after short vowels in medial position
bury, berry, very, cherish, flourish; worry, mirror, cherry

C18   that for /**j**/ they will use <y> before all vowels, or may leave it unspelt before /uː, ʊə, u/
yeast, yellow, yam, yacht, young, youth, yearn, year; unit, Euro, tenuous; pew, pure

C18a   that they will use <y> in medial position after a vowel, and <i> after a consonant
lawyer, loyal; million, stallion

C19   that for /**w**/ they will use <w> in all positions, except after /k/
will, wide; beware, reward; twin, dwindle, thwart, stalwart (but *quick, equate*)

Vowels
V1   that for /**ɪ, ɛ, æ, ɒ, ʌ**/ they will use <i, e, a, o, u> respectively
bid, bed, bad, cod, cud

V1a   that for /ʌ/ before /v/ they may use <o>
dove, plover, shovel

V2   that for /ʊ/ they will use <oo> before all consonants except /ʃ/, /l/, and <u> before /ʃ/, /l/
book, foot, push, full

V3   that for /iː/ they will use either double <e> or <ea> in all positions, but <i> before final /ʒ/ (<-ge>)
eel, eat; meet, meat; see, sea; prestige

V4   that for /ɑː/ they will use <ar> in all positions, but <a> before final /ʒ/ (<-ge>)
art; farm; car; collage

V5   that for /ɔː/ they will use either <or>, <oar>, <au> or <aw> in initial and medial positions
order, author, awkward; stork, caught, lawn

V5a   that they will use either <or>, <oar>, <oor>, <ore> or <aw> in final position
nor, door, more, paw (NB *awe, oar, ore*: 'three letter rule')

V6   that for /uː/ they will use <oo> in initial position
ooze

V6a   that they will use either <oo> or <u> in medial position
room, rumour

V6b   that they will use either <oo> or <ue> in final position
coo, glue

V6c   that for /juː/ they will use either <you> or <u> in initial position
you, youth, unit, unite

V6d   that they will use <u> in medial position and <ue> in final position for /juː/
fume, cue

V7   that for /ɜː/ they will use either <er>, <ear>, <ir> or <ur> in initial position
ermine, earn, earth; irk; urge

V7a   that they will use either <ear>, <er>, <ir> or <ur> in medial position
learn, serve, first, nurse

V7b   that they will use either <er>, <ir> or <ur> in final position
defer, fir, fur (NB *err*: 'three letter rule')

V8   that for /eɪ/ they will use either <a> or <ai> in initial and medial positions
acorn, aim; salient, whale, wail

V8a   that they will use <ay> in final position
say, essay

V9   that for /aɪ/ they will use <i> in initial position
idea, icon

V9a   that they will use <igh> or <i> before /t/ but only <i> before other consonants
light, like; sight; site

V9b   that they will use either <ie> or <y> for /aɪ/ in final position
pie, pry (NB *eye, buy, dye, rye*: 'three letter rule')

V10   that for /ɔɪ/ they will use <oi> in initial and medial positions, and <oy> in final position
oil, ointment; boil, voice; boy, alloy

V11   that for /əʊ/ they will use either <o> or <oa> in initial and medial positions
ozone, oath; lone, loan

V11a   that they will use either <oe> or <ow> in final position
roe, row (NB *owe*: 'three letter rule')

V12   that for /aʊ/ they will use either <ou> or <ow> in initial and medial positions, but only <ow> in final position
out, owl; cloud, crowd; cow, allow

V13   that for /ɪə/ they will use <ear> in all positions, but may use <eer> in initial and final positions, and also <ere> in final position
eerie, ear; beard; deer, dear, hear, here; volunteer

V14   that for /ɛə/ they will use <air> in all positions, but may use <are> and <ear> in final position, and <aire> in polysyllabic words in final position
air, airy, fairy, fair; stare, stair, bare, bear; solitaire

V15   that for /ʊə/ they will use either <oor>, <our> or <ure> in final position
poor, tour, dour, sure, contour

V15a   that they will use <ure> in final position for /jʊə/
pure, cure

V16   that for the weak vowel /**i**/ they will use <i> before a vowel, and <ie> after a solitary /k/ or otherwise <y> or <ey> in final position
piano, media; bookie, cookie, rookie; happy, lazy, money, donkey

V17   that for the weak vowel /**u**/ (and /**ju**/) they will use <u> before a vowel
situation, casual, tenuous

V18   that for the weak vowel /ə/ they will use <a> in initial position, any of <a, e, o, u, ou> in medial position, and either <a>, <er>, <re> or <ure> in final position
above, appeal; salary, celery, agony, century, jealous; sofa, comma, river, better, fire, centre, picture (NB *ire*: 'three letter rule')

V19   that they may adapt Vowel Hypotheses 1, 4, 5, 7 when /w/ precedes certain vowels as follows:
/w + ɒ/: <w + a> eg *wander, what, squat, swan*
/w + ʌ/: <w + o/u> eg *worry, wonder; swum*
/w + ɔː/: <w + ar/or>: eg *warn, worn, war, wore, swarm/sworn; ward, quart, dwarf, thwart, swarm* (but NB *qualm, quark*)
/w + ɜː/: <w + or>: eg *word, worse, worm, worth*

Additional grammatical information

Gr1   that for the perceived suffix **{ed}** for past forms of the verb, they will add <d> after a perceived <e> and <ed> after a consonant, and will change a final <y> which is preceded by a consonant to <i> and will double any other final solitary consonant which is preceded by a short vowel (unless that vowel is perceived to be spelt as a bigrapheme) – no matter what the actual pronunciation of the suffix might be
baked; added, passed, waited; tidied, applied; stopped, jogged; flooded

Gr2   that for the perceived suffix **{es}** for third person singular present forms of the verb and plural forms of nouns, they will add <es> after <s, sh, ch> and <o>, but will also double a final <z>, and <s> after any other vowel or other consonant, and will change a final <y> which is preceded by a consonant to <i> – no matter what the actual pronunciation of the suffix might be
passes, wishes, watches, potatoes, goes, does; quizzes; bakes, cakes, spas, commas; adds, stops, floods; puppies, rubies

Gr3   that for the perceived suffixes **{ing, ish, y, en, able}**, they will add <ing, ish, y, en, able> but will delete a final 'silent' <e> unless it marks a preceding <g> as /dʒ/,

and double a final solitary consonant which is preceded by a short vowel (unless that vowel is perceived to be spelt as a bigrapheme)
going, washing; yellowish; greeny, chewy; blacken, teachable
baking, caring; whitish, bluish; noisy; whiten; writable, manageable
winning, stopping; reddish; sloppy, gassy; redden; winnable

Gr4   that for the perceived suffix **{al}** for nouns derived from verbs and adjectives from nouns, they will add <al> but will delete a 'silent' <e> and will change a final <y> which is preceded by a consonant to <i>
arrival, refusal; virtual; denial, trial

Gr5   that for perceived **compound words**, they will either write the words together or separately, or may use a hyphen <-> if the first element ends in a vowel and the second begins with a vowel
teatime, tea time, pre-existent

Technical words
TC1   that for **/p, b, t, d/** they will use <p, b, t, d>, but without doubling

TC2   that for **/k/** they will use <c> before <a, o, u>, <ch> before <e, i, y>, and may use double <c> after a short vowel in medial position

TC2a that for **/kw/** they will use <qu>

TC2b that for /**ks**/ they will use either <x>, or double <c> before <e, i, y> in medial position; and in final position, either <x> for a substance, or <cs> for a study

TC3   that for /**g/** they will use **<g>**, but without doubling

TC4   that for /**m, n**/ they will use <m,n>, but without doubling

TC5   that for /**ŋ, ŋg**/ they will use <ng>

TC6   that for **/f/** they will use either <f> (but without doubling) or <ph>
pharmacy, philosophy, monograph

TC7   that for /**v**/ they will use Consonant Hypothesis 7

TC8   that for /θ/ they will use <th>

TC8a that they will not need to spell /ð/

TC9   that for /**s**/ they will use <s> in all positions, or may use <c> before <e, i, y>

TC10 that for /**z**/ they will use <z> (but without doubling), but may use <x> in initial position

TC11 that for /ʃ/ in medial position they will use Consonant Hypotheses 11a-d

TC12   that for /ʒ/ in medial position they will use Consonant Hypothesis 12

TC13   that for /**h**/ they will use Consonant Hypothesis 13

TC14   that for /**tʃ**/ they will use \<t\> before \<u\>

TC15   that they will not need to spell /**dʒ**/ in final position; in initial and medial positions, they will use \<g\> before \<e, i, y\>, and \<j\> otherwise (eg *conjunctivitis*)

TC16   that for /**l**/ they will use \<l\> in all positions, and may use doubling in medial position (eg *callisthenics*)

TC17   that for /**r**/ they will use either \<r\> or \<rh\> in initial or medial position, or double \<r\> or \<rrh\> in medial position (eg *diarrhoea*)

TC18   that for /**jʊ; jʊː, ju**/ they will either leave it unspelt before \<u\> or use \<eu\>

TC19   that they will not need to spell /**w**/, except after /k/; see **TC2a**

Vowels
TV1   that for /**ɪ, ɛ, æ, ɒ, ʌ**/ they will use Vowel Hypothesis 1, but may also use \<y\> for /ɪ/

TV2   that they will not need to spell /**ʊ**/

TV3   that for /**iː**/ they will use \<e\>

TV3a   that when /**iː**/ occurs in the final syllable with final /**n**/ they will use either \<ine\> or \<ene\>

TV3b   that when /**iː**/ occurs in the final syllable with any other consonant beside /n/, they will add \<e\>, unless they perceive a plural form (in which case they will not add \<e\>), eg *hypotheses*

TV4   that for /**ɑː**/ they will use Vowel Hypothesis 4

TV5   that for /**ɔː**/ they will use either \<or\> or \<au\> in initial and medial positions, but only \<or\> in final position

TV6   that for /**uː**/ they will use \<u\>

TV6a   that when /**uː**/ occurs in the final syllable, they will add \<e\>

TV7   that for /**ɜː**/ they will use either \<er\>, \<ir\> or \<ur\>

TV8   that for /**eɪ**/ they will use \<a\>

TV 8a that when /eɪ/ occurs in the final syllable, they will add <e>

TV9 that for /aɪ/ they will use either <i> or <y>

TV9a that when /aɪ/ occurs in the final syllable, they will add <e>

TV10 that for /ɔɪ/ they will use <oi>

TV11 that for /əʊ/ they will use <o>

TV11a that when /əʊ/ occurs in the final syllable, they will add <e>

TV12 that they will not need to spell /aʊ/

TV13 that for /ɪə/ they will use <ea> or <ia>

TV14 that for /ɛə/ they will use <a, ae> before /r/

TV15 that for /ʊə/ they will use <u, uo> before /r/

TV16 that for the weak vowel /i/ they will use <e> or <i> before a vowel and <y> in final position (eg *video-*, *audio-*, *-ology*)

TV17 that for the weak vowel /u/ they will use Vowel Hypothesis 17

TV18 that for the weak vowel /ə/ they will use Vowel Hypothesis 18

Scale of univocality

The scale of univocality measures the degree of bi-uniqueness of a grapheme, ie that one grapheme represents only one phoneme, and that one phoneme is represented only by that one grapheme (see Sgall, P 'Towards a Theory of Phonemic Orhtography' in Luelsdorff, P A (1987) *Orthography and Phonology* Amsterdam: John Benjamins). It is hypothesized that the orthographic competence of British educated young adults will exhibit the following degrees of univocality:

- absolute bi-uniqueness of graphemes <h, v, qu> and <ar>, ie they only represent /h, v, kw/ and /ɑː/, and that these phonemes are only represented by <h, v, qu> and <ar>
- relative bi-uniqueness of graphemes <b, d, f, g, l, m, n, p, r, z> which represent /b, d, f, g, l, m, n, p, r, z/, but these phonemes may be represented by a doubling of those graphemes
- relative bi-uniqueness of grapheme <ng> which represents /ŋ, ŋg/
- relative bi-uniqueness of grapheme <th> which represents /θ, ð/ and these phonemes are only represented by that grapheme
- relative bi-uniqueness of grapheme <j> which represents /dʒ/ but this phoneme is also represented by other graphemes
- relative bi-uniqueness of grapheme of <g> which represents /g/, and /dʒ/ before <e, i, y>

- relative bi-uniqueness of grapheme of <ge> which represents /dʒ/, and /ʒ/ in final position
- relative bi-uniqueness of grapheme of <w> which represents /w/ only, but /w/ is represented by <w> and <qu> in /kw/
- relative bi-uniqueness of grapheme of <x> which represents /ks/, but /k/ and /s/ are represented by other graphemes
- relative bi-uniqueness of grapheme of <k> which represents only /k/, but /k/ is represented by other graphemes
- relative bi-uniqueness of graphemes of <ai, ay> which represent /eɪ/
- relative bi-uniqueness of grapheme of <ea> which represents /iː/, but that phoneme is also represented by other graphemes
- relative bi-uniqueness of grapheme of <eer> which represents /ɪə/, but that phoneme is also represented by other graphemes
- relative bi-uniqueness of grapheme of <aire> which represents /ɛə/, but that phoneme is also represented by other graphemes
- regular deviation of the following graphemes:

  <c >: /k/, /s/ before <e, i, y>, /ʃ/ before <ia>, doubled for /ks/ before <e, i, y>

  <s >: /s, z/, doubled for /ʃ/ before –ion, –ure

  <t>: /t/, /ʃ/ before –ion, /tʃ/ before u, -ure

  <y>: /j/, /i/ in final position, and /aɪ/

  <a>: /æ, ei, ə, ɑː/, /ɪ/ in unstressed -age

  <e>: /ɛ, iː/ and 'silent', doubled for /iː/

  <i>: /ɪ, ɑi, i, iː/

  <o>: /ɒ, ʌ, əʊ, ə/, doubled for /ʊ, uː/

  <u>: /ʌ, ʊ, uː, ʊə, juː, jʊə, ju, ə/

  <au, aw, or, oar, oor, ore>: /ɔː/, <oor> also represents /ɪə/

  <ou, ow>: /aʊ/, but <ow> also represents /əʊ/ in final position

  <er, ir, ur>: /ɜː/; <er> also represents /ə/

  <ear>: /ɜː, ɪə, ɛə/; these phonemes are also represented by other graphemes

  <our>: /ɑʊə, ʊə/, but /ʊə/ is also represented by another grapheme

  <ure>: /ʊə, jʊə, ə/, but all these phonemes are represented also by other graphemes

P Tench, May 2007

1. kliː   clee/clea   Consonant Hypothesis 2; Vowel Hypothesis 3
2. rɪˈpɑː   repar   C1; V4
3. bəˈlɔː   belaw/belore   C16; V5a
4. kənˈvjuː   convue   V6d, 18
5. draʊ   drow   V12
6. veɪ   vay   C7; V8a
7. pəˈtrɜː   pertrir, -ur   V7b
8. kaɪ   kie, kye   General Hypothesis 6; C2; V9b
9. fɔɪ   foy   C6; V10
10. raɪə   rire   V9a, 18
11. trəʊ   trow, -oe   V11a
12. əˈnɪə   annear, -eer, -ere C4b; V13
13. bəˈdjʊə   bedure   V15a
14. raʊə   rower   V12, 18
15. əˈmɛə   ammare, -air –ear C4b; V14
16. klaɪt   clight, clite   V9a
17. gleɪt   glait, glate   V8
18. ˈkwɒtɪə   quattier C2g, 1a; V19, 16, 18
19. ˈsiːnɪŋ   seening, sea-, c- V3; C4; Grammar Hypothesis 3
20. ˈjuːflɪks   uflix, you-, -cks/cs C18, 2h; (or C2d; GrH 2)
21. tʃʌv   chove   C14, 7; V1a
22. ʌd   udd   General Hypothesis 6; C1a; V1
23. ˈhæzədʒ   hazage, -s-   V1, 18; C10a, 15b
24. ɪgˈzɜːstʃə   exersture, -i/ur- C3e, 14a; V7, 18
25. əˌpeɪʒəˈnɛə   appasionaire, -p- C1b, 12, 4; V18, 8, 14
26. ˈfræki   fracky, -ey   C2b, 16
27. ˈlɔːkɪʃ   laukish, -aw-, -or- C2; V5; Gr3
28. ˈsʌktʃuəl   suctual   C2, 14a, 16b; V1
29. ˈgærən   garren   C3, 17a; V1, 18 (or Gr3)
30. ˈpraʊnə   prouner, -ow-   C4; V12, 18
31. ˈkɒʃənət   cossionate   C2, 11a; V1, 18
32. brəʊð   brothe, -oa-   C8a; V11
33. dʊp   doop   C1; V2
34. ˈdʒɪpəl   jipple, g-   C15, 1a, 16b
35. wəʊθ   woth, -oa-   C19, 8a; V11
36. flɪdʒ verb
37. ˈhaɪdrʌm adjective
38. ˈeɪzəpriːn noun
39. sɔːˈkeɪʃən noun (a process)
40. daɪˈlʌkʃuəs adjective
41. ˈvɔɪsəʊklaɪd noun
42. nuːˈmɛtʃərə adjective
43. ˈzɪərɒks noun
44. ruːˈmætsɔɪd noun
45. zaɪˈlɒdʒɪnəs adjective
46. glɪθɪˈrɔɪtɪks noun
47. krɔɪtɪˈlaɪtəs noun
48. θaɪləˈdʒɪlɪks noun
49. gləˈmætsəˈrɪə noun

50. ˈdʒɛərɔɪˈsɛtəliːn noun