



Defnyddio Barn Gymharol Addasedig ar gyfer asesu ymatebion Asesiadau Di-arholiad (NEA) TGAU Hanes

Adroddiad Ymchwil

Mehefin 2022





Defnyddio Barn Gymharol Addasedig ar gyfer asesu ymatebion i NEA TGAU Hanes. Adroddiad Ymchwil.

Awdur: Vasile Rotaru, Uwch Swyddog Ymchwil Cymwysterau Cymru

Cydnabyddiaethau

Hoffwn ddiolch i Tom Anderson (Pennaeth Ymchwil ac Ystadegau Cymwysterau Cymru) am ei gyfraniad sylweddol at gysyniad a chynllun yr ymchwil, a hefyd am ei gymorth i adolygu canfyddiadau a drafftiau dilynol yr adroddiad hwn yn feirniadol.

Hoffwn ddiolch i ddau aelod o Grŵp Ymchwil a Chynghori CC - Dr Joshua McGrain (Athro Cysylltiol mewn Asesu Addysgol yn Rhydychen) ac Anne Pinot de Moira (Cymrawd Anrhydeddus Norham ym Mhrifysgol Rhydychen, yr Adran Addysg) - am sylwadau ac awgrymiadau a wellodd yr adroddiad yn fawr.

Rwyf yn ddiolchgar iawn i Mark Hogan am ei gymorth gyda'm cwestiynau ar yr agweddau ystadegol.

Diolch hefyd i'r athrawon a gymerodd ran yn yr ymchwil hon am eu parodrwydd i gyfrannu ac am eu trafodaethau agored.

Hoffwn ddiolch i RM Education am ddarparu mynediad am ddim i feddalwedd RM Compare™.

Nid yw'r cydnabyddiaethau'n awgrymu cymeradwyo'r astudiaeth gyfredol a'i chanfyddiadau.

Cyhoeddwyd ym mis Mawrth 2022 gan Cymwysterau Cymru

© 2022 Cymwysterau Cymru

I ddyfynnu'r cyhoeddiad hwn: Rotaru, V. (2022) Defnyddio Barn Gymharol Addasedig ar gyfer asesu ymatebion i NEA TGAU Hanes. Cymwysterau Cymru

Cynnwys

1. Cyflwyniad	3
2. Cefndir ymchwil	3
3. Nodau ymchwil	6
4. Dull	7
4.1. Beirniaid.....	7
4.2. Sgriptiau.....	7
4.3. Gweithdrefn.....	10
4.4. Cyfweiliadau ar ôl y gweithdy.....	11
5. Canlyniadau Barn Gymharol	12
6. Profiad cyfranogwyr o BG a'r agweddau a ystyriwyd ganddynt wrth feirniadu	17
6.1. Dulliau o ddarllen ymatebion.....	18
6.2. Yr angen i ddarllen ymatebion yn llawn.....	19
6.3. Agweddau a ystyriwyd wrth wneud y penderfyniadau.....	21
6.3.1. Gwerthuso a dadansoddi dibynadwyedd a defnyddioldeb y ffynonellau....	25
6.3.2. Gwybodaeth.....	25
6.3.3. Cymhlethdod y dadansoddiad.....	26
6.3.4. Arddull ysgrifennu fecanistig.....	26
6.3.5. Rôl yr elfennau strwythurol.....	27
6.3.6. Pwysigrwydd y geiriau allweddol.....	28
6.4. Agweddau heriol ar feirniadu.....	29
6.4.1. Beirniadu ymatebion o werth tebyg.....	29
6.4.2. Beirniadu ymatebion ar wahanol bynciau.....	30
6.4.3. Beirniadu ymatebion ysgrifenedig.....	32
6.4.4. Gwneud sylwadau.....	33
6.5. Cwestiwn BG.....	36
7. Manteision canfyddedig BG	36
7.1. Effeithiau ar yr asesiad.....	36
7.1.1. Effeithiau ar ddilysrwydd a chreadigrwydd yr ymatebion.....	36
7.1.2. Tegwch.....	37

7.1.3.	Rhwyddineb gwneud penderfyniadau	38
7.1.4.	Cyfleustra	40
7.1.5.	Baich gweinyddol	41
7.2.	Yr effaith ar ddatblygiad proffesiynol	42
8.	Gweithredu BG fel dewis yn lle marcio confensiynol	43
8.1.	Cynnwys a hysbysu athrawon	44
8.2.	Canllawiau BG i athrawon a dysgwyr	45
8.3.	Hyfforddiant	46
8.4.	Hyd yr ymatebion	47
8.5.	Mathau o ymateb.....	48
8.6.	Yr amser sydd ei angen.....	48
9.	Cyfyngiadau astudio	51
10.	Casgliadau ac ymchwil bellach	51
	Cyfeiriadau	56

1. Cyflwyniad

Dull yw Barn Gymharol (BG) lle mae grŵp o feirniaid yn cymharu parau o symbyliadau a dynnwyd o'r un set dro ar ôl tro (er enghraifft, parau o draethodau a ysgrifennwyd gan y dysgwyr o'r un dosbarth). Ar gyfer pob pâr, byddai'r beirniaid yn dewis enillydd yn seiliedig ar unrhyw 'briodoledd meintiol neu ansoddol y gallwn ei ystyried yn "fwy" neu "lai" ar gyfer pob sbesimen' (Thurstone, 1927, t. 273), a allai hefyd olygu i raddau "gwell", "mwy", "mwy deniadol" ac ati.

Nod ymchwil Cymwysterau Cymru oedd archwilio barn athrawon Hanes am BG fel dull yn lle marcio traddodiadol ac ymchwilio i'w penderfyniadau wrth feirniadu. Felly, astudiaeth ansoddol yw hon yn bennaf er gallai'r canlyniadau BG ychwanegu at y canfyddiadau presennol ar ddibynadwyedd a dilysrwydd BG.

2. Cefndir ymchwil

Mae BG yn seiliedig ar gyfraith barn gymharol a ddatblygwyd gan y seicolegydd Louis Thurstone (Thurstone, 1927). Tybiodd bob tro y bydd person yn cymharu pâr o symbyliadau fod y ddau symbyliad yn cael eu gosod ar raddfa neu gontinwmm drwy 'broses wahaniaethu'. Mae hon yn 'fath o broses ynom lle rydym yn ymateb yn wahanol' i symbyliadau ac yn eu categoreiddio yn ôl graddau o feddu ar briodoledd penodol. Drwy roi'r symbyliad ar y continwmm, rydym yn aseinio gwerth ar unwaith iddo. Nid union werth mo hwn, ond drwy gymharu symbyliad dro ar ôl tro rydym yn cael set o werthoedd sy'n ffurfio dosbarthiad arferol. Gwerth cymedrig y dosbarthiad yw bras werth graddfa'r symbyliad (Thurstone, 1927). Mae gwerthoedd pob symbyliad ar raddfa cyfyngau lle cymerir bod trefn restrol y symbyliadau a'r gwahaniaethau rhwng sgoriau yn ystyrllon. Defnyddir y gwerth hwn i roi trefn restrol ar symbyliad mewn perthynas â symbyliadau eraill.

Mae'r ymchwil gyfredol yn canolbwyntio ar ddefnyddio math o BG mewn asesu addysgol, lle gofynnir i grŵp o feirniaid gymharu parau o wrthrychau a dynnwyd o'r un set. Ceir dulliau yn lle defnyddio BG neu ddefnyddio cymariaethau hefyd. Er enghraifft, gellir gofyn i feirniaid gymharu a rhoi cyfres o ymatebion mewn trefn restrol, yn hytrach na chymharu'r parau (gweler Bramley, 2005). Gellir defnyddio BG hefyd i gynllunio asesiad dau gam lle mae grŵp o feirniaid yn graddnodi cyfres o ymatebion i ddechrau, ac wedyn defnyddir ymatebion dethol o'r set hon fel pwyntiau angori ar raddfa, a marcwyr yn cymharu'r ymatebion mewn perthynas â hwy (McGrane et al., 2018).

I weithredu BG, mae'n ofynnol i feirniaid gymharu digon o barau i benderfynu'n ddigon dibynadwy werthoedd y symbyliadau (Thurstone, 1927). Gellid dewis y parau

ar hap neu drwy ddefnyddio dull addasedig. Nid yw Barn Gymharol Addasedig (BGA), a ddefnyddir yn yr astudiaeth hon, yn paru gwrthrychau sydd ymhellach ar wahân (hynny yw, gwrthrych sydd bron yn sicr â mwy o werth paramedr na'r llall), gan nad yw penderfyniad ar bâr o'r fath yn ychwanegu gwybodaeth ddefnyddiol at y diben graddio. Gan nad oes gennym unrhyw wybodaeth am werth gwrthrych y rhan fwyaf o'r amser, caiff y parau eu dewis ar hap i ddechrau. Pan fydd y broses feirniadu yn darparu rhywfaint o wybodaeth am werth y gwrthrychau, mae algorithm yn llywio'r broses o ddethol parau. Gyda phob cylch o'r detholiad addasedig hwn, gwahanir gwrthrychau'n well i'r graddau y gellir cyflawni canlyniadau dibynadwy gyda nifer gymharol is o gymariaethau (gweler Pollitt, 2012), ac mewn tua deg cylch yn gynharach nag yn y detholiad o barau nad yw'n addasedig (ar hap yn unig) (Kimbell, 2021).

Yn ôl Marshal et al. (2020) ymchwiliwyd fwyfwy i ddefnyddio BG mewn asesu addysgol. Mae BG wedi'i harchwilio, ei threalu a'i defnyddio mewn asesu (ar dasgau ffurfiannol a chrynodol) ac wrth gynnal safonau wrth ddyfarnu graddau cymhwyster.

Awgryma'r canfyddiadau fod gan BG, ar gyfer mathau penodol o asesiadau, rywfaint o botensial fel dewis yn lle marcio confensiynol a chymedroli allanol. Dadleuodd Tarricone a Newhouse (2016, t.13) hyd yn oed fod BG yn 'ddewis dibynadwy iawn yn lle marcio dadansoddol traddodiadol' a allai wella dilysrwydd a dibynadwyedd arholiadau traddodiadol lle mae llawer yn y fantol.

O ran dilysrwydd, mewn sgiliau crefft ymarferol (technegau prosesu addurniadol pren a metel) dangoswyd bod defnyddio BG yn rhoi mwy o ryddid i ddysgwyr greu (drwy beidio â'i gwneud yn ofynnol iddynt ddilyn meini prawf asesu caeth) ac yn gwella gallu meini prawf asesu i newid mewn ymateb i sut mae myfyrwyr yn cysoni eu gwaith â thystiolaeth o allu (Seery et al., 2012). Yn yr un modd, dangoswyd bod dull asesu BG, mewn mathemateg, yn annog cynllunio tasgau llai strwythuredig a mwy seiliedig ar broblemau o'u cymharu â'r tasgau nodweddiadol sy'n seiliedig ar gynllun marcio (Jones ac Inglis, 2015).

Ffactor sydd wedi gwneud i BG ennyn diddordeb oedd y canfyddiadau ei bod yn fwy dibynadwy na marcio traddodiadol (Kimbell, 2021). Er enghraifft, canfu trosolwg gan Bramley (2015, t.6) fod naw cyhoeddiad yn disgrifio astudiaethau a oedd â 9 i 15 o gymariaethau fesul sgrïpt, yn dangos lefelau dibynadwyedd a oedd yn amrywio o 0.86 i 0.97, gydag un eithriad o 0.73. Roedd yr astudiaethau hyn yn cynnwys pynciau fel Mathemateg, Cemeg, Dylunio, a gwahanol fathau o ysgrifennu.

Fodd bynnag, nid yw'r astudiaethau'n argyhoeddiadol ynghylch manteision posibl BG. Er enghraifft, cwestiynwyd dilysrwydd BG wrth asesu gwahanol gymwyseddau, oherwydd ni allai un farn gyfannol grynhoi 'natur anwastad perfformiadau ymgeiswyr' (Barkaoui, 2011 a ddyfynnwyd yn Whitehouse a Pollitt, 2012). Yn yr un

modd, er bod asesiad BG fel arfer â mynegeion dibynadwyedd uchel, nodwyd bod ymaddasolrwydd yn tueddu i oramcangyfrif y cyfernod dibynadwyedd (Bramley 2015). Awgrymwyd y gellir unioni'r broblem hon drwy fireinio'r algorithm dethol, gan sicrhau hefyd fod tua 15 o gylchoedd beirniadu yn cael eu cynnal (Kimbell, 2021).

Mae'r rhan fwyaf o'r ymchwil bresennol yn gwerthuso rhinweddau asesiad BG gan ddefnyddio canlyniadau beirniadu (e.e. cyfernodau cydberthyniad/dibynadwyedd, amser beirniadu cyfartalog). Mae rhai astudiaethau'n archwilio barn beirniaid am y dull neu eu dull o feirniadu. Mewn rhai ohonynt bu'r ymchwilwyr yn arolygu beirniaid (Jones ac Inglis, 2015; Marshal et al., 2020; Pollitt, 2012), ac mewn eraill buont yn archwilio sylwadau a adawodd beirniaid am y sgriptiau neu eu barnau (Daal, 2016; Jones et al., 2015; Whitehouse, 2013). Hyd y gwyddom, dim ond dwy astudiaeth (Coertjens et al., 2021 a Newhouse, 2014) sydd wedi archwilio'n ansoddol farn beirniaid. Ni fyddai pob un o'r beirniaid yn yr astudiaethau hyn yn asesu'n gonfensiynol yn rheolaidd (neu o gwbl) ymatebion tebyg i'r rhai a farnwyd.

Fodd bynnag, os yw BG am gael ei defnyddio wrth asesu, credwn fod angen ymchwiliad manylach i farn aseswyr i archwilio sut y byddai BG yn cyd-fynd â pharadeim asesu penodol a sut y byddai'n mynd i'r afael â'r cyfyngiadau y mae aseswyr yn eu hwynebu.

Gan fod defnyddio BG yn awgrymu newid y dull asesu, mae agwedd ddefnyddiol i'w harchwilio yn ymwneud ag agoredrwydd marcwyr i dderbyn BG a'r ffactorau a allai helpu, neu lesteirio, eu cymeradwyaeth. Wedi'i gysylltu'n agos â hyn hefyd mae'r angen i archwilio pa fath o gymorth parhaus y gallai fod ei angen ar feirniaid wrth ddefnyddio BG wrth asesu. Un agwedd arall sy'n werth ei harchwilio yw a ydy proses gwneud penderfyniadau BG yn caniatáu i feirniaid dalu sylw ac asesu'r nodweddion sy'n cael eu hasesu drwy farcio confensiynol. Gallai canfyddiadau marcwyr o'r dull fod yn ddefnyddiol wrth archwilio'r agweddau hyn.

Yn ogystal, byddai'n ddefnyddiol ymchwilio i weld a allai BG fod yn offeryn asesu hyfyw, neu well hyd yn oed, o ran hydrinedd, dilysrwydd a dibynadwyedd na marcio confensiynol ar gyfer asesiad crynodol o ymatebion ysgrifennu estynedig. Mae'r system gymwysterau gyffredinol yng Nghymru yn dibynnu ar ymatebion adeiledig, gan gynnwys eitemau ymateb estynedig (y mae rhai ohonynt wedi'u marcio gan athrawon yn fewnol), yn hytrach na defnyddio cwestiynau aml-ddewis neu gwestiynau ymateb caeedig. Credir yn aml fod dull gweithredu o'r fath yn sicrhau mwy o ddilysrwydd yr asesiad (gweler, er enghraifft, Murphy ac Yancey, 2009; Kimbell, 2021). Fodd bynnag, yn achos ysgrifennu estynedig, mae ymchwil yn awgrymu y gall cywirdeb marcio cymharol isel effeithio ar yr asesiad. Er enghraifft, mae Rhead et al. (2018, t.4) wedi dangos bod y tebygolrwydd o gael y 'radd ddiffiniol'¹ yn amrywio'n

¹ Yn yr astudiaeth a ddyfynnwyd, mae'r term 'diffiniol' yn cyfeirio at y marc a roddwyd gan banel o uwch arholwyr ar lefel eitem ar gyfer pob ymateb dosbarthu.

sylweddol ar draws cymwysterau yn Lloegr. Roedd y cysondeb marcio ar ei isaf ar gyfer arholiadau a oedd yn cynnwys atebion ysgrifennu estynedig: 0.52 ar gyfer cymwysterau iaith a llenyddiaeth Saesneg, gyda hanes yn cael yr ail gysondeb isaf. Ar y llaw arall, dibynadwyedd y marcio ar gyfer mathemateg oedd 0.92. Mewn astudiaeth arall, roedd dibynadwyedd marcio cwestiynau ysgrifennu arholiad yn seiliedig ar ffynhonnell UG Hanes yn amrywio o 0.52 i 0.62 (Holmes et al., 2018, t.17).

Mae'n debygol bod asesu a arweinir gan athrawon, fel y'i cyferbynnir â marcio allanol, yn cael effaith bellach ar gywirdeb marcio oherwydd yr agweddau ychwanegol y gallai athrawon eu hwynebu neu eu hystyried wrth farcio (e.e. rhagfarnau a ddatblygwyd am fod yr athrawon yn adnabod y myfyrwr, neu oherwydd bod canlyniadau cymwysterau'n cael eu defnyddio mewn prosesau atebolrwydd ysgolion) (gweler er enghraifft, Kellaghan et al., 2019, t.263-264).

Cafwyd ymchwil a ddangosodd y gallai BG fod yn ddull dilys ar gyfer tasgau ysgrifennu TGAU neu Safon Uwch (Bramley a Vitello, 2019; Steedle a Ferrara, 2016; Whitehouse a Pollitt, 2012; Pollitt a Crisp, 2004)². Er enghraifft, dangosodd Whitehouse (2013) fod beirniaid sy'n asesu sgriptiau Daearyddiaeth safon UG yn defnyddio'r cynllun marcio a'r amcanion asesu wrth farnu, hyd yn oed os na ddarperir y cynllun marcio iddynt. Mae'r astudiaethau hyn wedi defnyddio ymatebion gwirioneddol neu wedi'u haddasu ar gyfer eitemau arholiadau, ond ym mhob un ohonynt roedd y sgriptiau'n gymharol fyr. Fodd bynnag, gallai eitemau ymateb ysgrifennu estynedig fod ag o leiaf dair tudalen o lawysgrifen. Er enghraifft, dylai ymateb TAG Safon UG/Safon Uwch Hanes fod rhwng 3,000 a 4,000 o eiriau; dylid cyflwyno Prosiect Unigol Bagloriaeth Cymru Genedlaethol/Sylfaen CA4 ar ffurf ysgrifenedig rhwng 1,000 a 2,000 o eiriau.

Gallai sgriptiau hirfaith ei gwneud yn anoddach ffurfio barn, a allai effeithio ar effeithlonrwydd BG gan fod yn rhaid ystyried mwy o gynnwys i lunio barn. Er enghraifft, asesodd Jones et al. (2015) bapur TGAU Mathemateg cyfan gan ddefnyddio BG, a rhai o'r sgriptiau hyd at 47 tudalen. Canfuwyd bod hyd sgriptiau (ac, yn ein barn ni, natur yr ymatebion hyn mae'n debyg) yn ei gwneud yn anos llunio barn gyffredinol ac yn gwneud beirniadu'n anodd, yn straen ac yn llai effeithlon.

3. Nodau ymchwil

Nodau'r astudiaeth hon oedd:

- archwilio barn athrawon am ddefnyddio BG i farcio ymatebion ysgrifennu estynedig (a ddiffinnir fel ymatebion ag o leiaf dair tudalen o lawysgrifen) gan gynnwys gwybodaeth am sut y lluniwyd y barnau;

² Sylwer bod rhai astudiaethau wedi defnyddio ymatebion estynedig i archwilio defnyddio BG wrth gynnwys safonau (gweler yr enghreifftiau a nodwyd yn Steedle a Ferrara, 2016).

- casglu gwybodaeth sy'n berthnasol i ddilysrwydd, dibynadwyedd a hydrinedd (e.e. yr amser sydd ei angen i farnu o'i gymharu â marcio) defnyddio BG ar gyfer asesu crynodol atebion estynedig.

4. Dull

4.1. Beirniaid

Recriwtiwyd 22 o athrawon hanes yn Ionawr 2020 a oedd wedi marcio naill ai TGAU Hanes neu Safon Uwch Hanes i gymryd rhan mewn gweithdy BG drwy wahoddiad a anfonwyd drwy lythyr i bob ysgol yng Nghymru. Nid oedd yr un o'r cyfranogwyr yn gwybod am BG cyn y gweithdy. Cynhaliwyd y gweithdy yng Nghaerdydd ac roedd y rhan fwyaf o'r athrawon yn dod o dde a de-orllewin Cymru. Fodd bynnag, fel y gwelir yn Nhabl 1, roedd gan y cyfranogwyr ystod dda o brofiad marcio (TGAU neu Safon Uwch) ac roeddent yn cynrychioli'r prif fathau o ganolfan yng Nghymru.

Tabl 1: Profiad marcio athrawon a'r math o ysgol

Math o ysgol	Blynyddoedd o brofiad				Cyfanswm
	1-2	3-4	5-6	7+	
A Gynhelir	-	1	1	12	14
Annibynnol	2	2	-	1	5
Coleg Addysg Bellach	1	1	1		3
Cyfanswm	3	4	2	13	22

4.2. Sgriptiau

Barnodd y cyfranogwyr 23 o ymatebion rhan (a) Asesiad Di-Arholiad (NEA) TGAU Hanes a oedd wedi'u marcio'n gonfensiynol yn 2019.

Mae NEA yn elfen orfodol o TGAU Hanes. Mae'n ategu'r arholiadau allanol ac mae'n werth 40 marc, sef 20% o'r cymhwyster TGAU Hanes. Mae'r NEA wedi'i gwblhau yn cynnwys dau ddarn o waith ysgrifenedig sydd wedi'u marcio ar wahân. Mae tasg rhan (a) yn gofyn i ymgeiswyr ddadansoddi a gwerthuso ffynonellau hanesyddol sy'n mynd i'r afael â'r maes pwnc a ddewiswyd ac mae'n werth 14 marc. Mae tasg rhan (b) yn gofyn i ymgeiswyr drafod mater o ddadl hanesyddol sy'n codi o'r maes pwnc a ddewiswyd ac mae'n werth 26 marc.

Mae pob canolfan yn dewis y maes pwnc a fydd yn cael ei drafod wrth gwblhau uned NEA. Asesir yr ymatebion rhan (a) ar draws tri Amcan Asesu:

- AA1 - Dangos gwybodaeth a dealltwriaeth o nodweddion allweddol y cyfnodau a astudiwyd (2 farc)
- AA2 - Esbonio a dadansoddi digwyddiadau a chyfnodau hanesyddol a astudiwyd gan ddefnyddio cysyniadau hanesyddol gradd dau (4 marc)

- AA3 - Dadansoddi, gwerthuso a defnyddio ffynonellau i lunio barnau a brofwyd, yng nghyd-destun y digwyddiadau hanesyddol a astudiwyd (8 marc).

Mae tasgau'r NEA yn cynnwys pecyn o ddogfennau sy'n cynnwys hyd at 25 o ffynonellau (ffotograffau, cartwnau, areithiau, ac ati). Yn y pecyn ceir ffynonellau cyfoes a diweddarach, sy'n amlinellu o leiaf ddau ddehongliad o'r pwnc. Mae ymgeiswyr yn defnyddio'r pecyn ffynhonnell yn sail i'w hymchwil 8 i 10 wythnos, sy'n cael ei chynnal o dan oruchwyliaeth gyfyngedig. Mae tasg rhan (a) yr NEA wedyn yn gofyn i fyfyrwyr ddewis pedair ffynhonnell a dadansoddi pa mor ddefnyddiol a dibynadwy ydynt wrth esbonio agwedd benodol ar y pwnc a astudiwyd. Rhaid i ddysgwyr gwblhau'r asesiad rheoledig cyfan o fewn pum awr o amser ffurfiol dan oruchwyliaeth, a dylai rhan (a) gymryd hyd at ddwy awr i'w chwblhau. Gall ymgeiswyr ysgrifennu neu deipio'r ymatebion. Cynghorir y dylai ymatebion rhan (a) fod tua 1,000 o eiriau, ond nid yw hwn yn derfyn gorfodol.

Pan fydd ymateb rhan (a) wedi'i gwblhau, bydd athrawon yn ei farcio ac yn cymedroli'r marcio'n fewnol, ac yna bydd y corff dyfarnu yn dewis nifer penodol o ymatebion ar hap i'w cymedroli'n allanol.

Ar gyfer y gweithdy BG, gofynnwyd i dri chyfranogwr – o ysgol gyfun cyfrwng Saesneg, ysgol ddwyieithog (cyfrwng Cymraeg/Saesneg) ac ysgol annibynnol – gyflwyno ymatebion gan ddsbarth cyfan drwy wefan ddiogel. Roedd yr ymatebion gan ddwy ysgol yn mynd i'r afael â pha mor ddefnyddiol a dibynadwy oedd y ffynonellau a ddewiswyd wrth ddangos polisi tramor John F. Kennedy. Archwiliodd yr ymatebion gan yr ysgol arall pa mor ddefnyddiol a dibynadwy oedd y ffynonellau o ran dangos pa mor anodd oedd bywyd i drigolion Dwyrain Llundain yn ail ran y 19eg ganrif.

Cyflwynwyd cyfanswm o 67 o ymatebion, ac roedd 34 ohonynt wedi'u hysgrifennu â llaw. Roedd ymatebion o un ysgol i gyd wedi'u teipio ac roedd nifer fach o'r ymatebion gan y ddwy ysgol arall hefyd wedi'u teipio. Roedd y rhan fwyaf o'r ymatebion wedi'u teipio rhwng 1,500 a 2,000 o eiriau o hyd; roedd cyfran lawer llai dros 2,000 o eiriau; a dim ond ychydig oedd tua 1,200 o eiriau. Roedd bron pob un o'r ymatebion hyn yn dair neu bedair tudalen. Roedd hyd yr ymatebion llawysgrifen yn amrywio o bump i ddeg tudalen. Roedd gan bron i hanner ohonynt wyth tudalen, ac roedd gan tua chwarter chwe thudalen.

Yr ystod o farciau sydd ar gael ar gyfer ymateb rhan (a) yw 0 i 14. Roedd dosbarthiad y marciau yn y sampl a gyflwynwyd yn amrywio o 7 i 14 gyda chanolrif o 11, felly nid oedd yn cynrychioli'r ystod marciau lawn a oedd ar gael ar y dasg. Ar draws y tair ysgol, roedd y marc cyfartalog yn amrywio o 10.7 i 12.1.

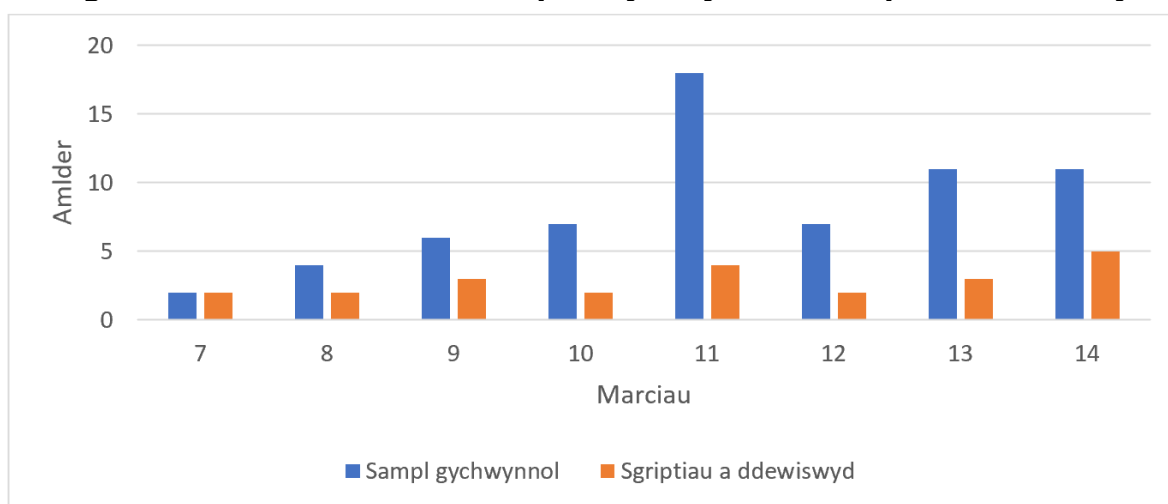
Er mwyn cyrraedd nifer dderbyniol o farnau ar gyfer pob ymateb, dim ond 23 o ymatebion, sgrïptiau o hyn ymlaen, a ddewiswyd ar gyfer yr ymarfer BG, ac roedd 11 ohonynt wedi'u hysgrifennu â llaw (Tabl 2).

Tabl 2: Sampl cychwynnol o ymatebion a sgrïptiau a ddewiswyd

	Sampl cychwynnol	Sgrïptiau a ddewiswyd
Cyfanswm	67	23
O'r rhain, mewn llawysgrifen	34	11
Marc canolrif	11	11
Marc cymedrig	11.4	11.04
Ystod o farciau	7-14	7-14
Gwriad safonol o farciau	1.93	2.36

Dewiswyd y sgrïptiau ar hap ond ein nod oedd cael sampl gynrychioliadol ar draws pob grŵp marcio ym mhob dosbarth. Anelwyd hefyd at sicrhau bod ymatebion gyda'r marciau isaf yn cael eu gorgynrychioli, i gyfrif am nifer fach ar y gwerthoedd marcio isaf. Mae dosbarthiad marciau'r sgrïptiau a ddewiswyd ychydig yn wahanol i'r sampl gychwynnol, er bod y marc canolrif yr un fath ag yn y sampl gychwynnol (11), tra bod y cyfartaledd yn agos at y sampl gychwynnol – 11.04 gyda gwriad safonol uwch o 2.36 (y gellir ei briodoli i orsamplu sgrïptiau sgorio is). Roedd hyd y sgrïptiau wedi'u teipio yn amrywio o 1,147 i 1,810 o eiriau, ac roedd gan y sgrïptiau llawysgrifen ddosbarthiad tebyg o nifer y tudalennau i'r sampl gychwynnol.

Ffigur 1: Dosbarthu marciau samplau cychwynnol a samplau a ddewiswyd



Roedd yr holl sgrïptiau'n ddiennw, a dilëwyd yr anodiadau a'r marciau presennol i sicrhau nad oeddent yn effeithio ar farn y beirniaid.

4.3. Gweithdrefn

Mynychodd y cyfranogwyr weithdy undydd. Yn ystod rhan gyntaf y gweithdy, dysgon nhw am BG a chynhaliwyd sesiwn feirniadu brawf lle'r oeddent yn cymharu darllenadwyaeth llawysgrifen yr ymatebion a ddewiswyd. Roedd hwn yn ymarfer hyfforddi ond roedd hefyd yn darparu gwybodaeth a oedd yn caniatáu i ni asesu a oedd llawysgrifen yn cyflwyno rhagfarn i farnu ansawdd (gweler adran 5). Cynhaliwyd cyfanswm o 20 o gylchoedd beirniadu, gan gyflawni cyfernod dibynadwyedd o 0.82.

Beirniadwyd yr ymatebion yn ail ran y gweithdy a pharhaodd y gwaith hwn am ryw bedair awr. Cafodd y cwestiwn y gofynnwyd i'r beirniaid ei ystyried ei lunio yn yr un modd â'r cwestiwn yn Pollitt a Crisp (2004): *Pa sgrïpt sy'n cynrychioli disgybl sy'n well am Hanes?*

Roeddem wedi rhoi gwybod i'r cyfranogwyr am bynciau'r ymatebion ac wedi anfon y ffynonellau perthnasol atynt cyn y gweithdy. Ar y diwrnod, rhoesom y cynllun marcio iddynt a chyfarwyddo y gallent ei ddefnyddio fel canllaw wrth benderfynu ar yr enillydd ond nad oedd hyn yn orfodol. Rhoddwyd cyfarwyddyd i'r cyfranogwyr lunio barn yn gyfleus o gyflym ac i beidio â phoeni am yr angen i gwblhau nifer penodol o farnau neu boeni am fod y tu ôl i eraill wrth gwblhau'r dasg. Fe'u cyfarwyddwyd i lunio'r farn pryd bynnag yr oeddent yn teimlo eu bod yn barod i wneud hynny.

Rhannwyd y cyfranogwyr yn ddau grŵp cyfartal gan ddefnyddio haenau hap i sicrhau dosbarthiad tebyg o athrawon gyda'r un profiad marcio. Gofynnwyd i un grŵp (Grŵp F o hyn ymlaen) farnu'r ymatebion a rhoi adborth neu sylwadau ar bob ymateb. Cyfarwyddwyd y gallai'r sylwadau fod yn frawddegau byr neu hyd yn oed grŵp o eiriau a fyddai'n disgrifio'r ymateb ac nad oedd angen i'r sylwadau fod yn debyg i'r anodi a wnaed wrth farcio. Gallai cyfranogwyr weld ac ychwanegu at eu sylwadau wrth ddod ar draws sgrïpt, y tro nesaf, ond ni allent weld y sylwadau a wnaed gan feirniaid eraill. Roedd y grŵp arall (Grŵp NF o hyn ymlaen) yn beirniadu'r un ymatebion ond nid oedd yn rhaid iddo ddarparu unrhyw adborth/sylwadau.

Lluniwyd barnau gan ddefnyddio meddalwedd RM Compare™. Mae'n cyflwyno ymatebion parau o fyfyrwyr ar-lein drwy borwr rhyngwyrdd ac mae'r beirniad yn dewis naill ai'r ymateb ar y chwith neu'r dde drwy glicio ar y fysell berthnasol a neilltuwyd ar y bysellfwrdd.

Yn y cylch beirniadu cyntaf, mae'r feddalwedd yn dewis y parau ar hap. Yn y tri chylch nesaf, mae'n paru'r ymatebion gan ddefnyddio dull Twrnamaint y Swistir. O dan y dull hwn, yn y cyntaf allan o'r tri chylch, mae'r enillwyr a'r collwyr yn cael eu paru yn y drefn honno gydag enillwyr a chollwyr, ac yn y ddau gylch nesaf, mae'r dull yn paru'r sgrïptiau sydd â'r un nifer o enillion. Ar ôl cwblhau'r pedwar cylch, mae'r meddalwedd yn paru'r sgrïptiau gan ddefnyddio algorithm. Mae'r algorithm yn dewis y sgrïptiau sydd â'r gwerthoedd paramedr ansawdd agosaf, a amcangyfrifir o'r

newydd ym mhob cylch, yn seiliedig ar y farn a luniwyd gan bob beirniad (am ddisgrifiad manylach gweler Rangel-Smith a Lynch, 2018). Felly, po fwyaf o gylchoedd a gynhelir, yr anoddaf y dylai fod i lunio'r farn, gan y dylai'r sgriptiau pâr fod â gwerthoedd cynyddol agosach. Mae dewis sgriptiau yn seiliedig ar ganlyniadau barnau blaenorol yn gwneud y barnau'n addasedig, felly gelwir y dull yn BG Addasedig (Bramley a Vitello, 2019). Yn ein hachos ni, roedd yr ymatebion hefyd wedi'u cadwyno, h.y. yn ystod y cyfnod addasedig, dewisodd y feddalwedd un ymateb o'r farn flaenorol i fod yn rhan o'r pâr nesaf.

Ni chafodd cyfranogwyr unrhyw anawsterau amlwg wrth lywio'r feddalwedd BG a daethant i arfer â hi'n gyflym. Yr unig anghyfleustra a grybwyllwyd gan lawer oedd yr angen i sgorlio i fyny'r dudalen i bwysu'r botwm 'tudalen nesaf'. Credai'r cyfranogwyr y byddai gosod y botwm hwn ar waelod y dudalen yn gwneud y rhyngwyneb yn haws ei ddefnyddio.

4.4. Cyfweliadau ar ôl y gweithdy

Ar ôl y gweithdy, cynhaliwyd cyfweliadau lled-strwythuredig â phob cyfranogwr. O'r 22 o gyfranogwyr, cyfwelwyd â 10 yn ystod yr wythnos gyntaf ar ôl y gweithdy, a'r gweddill yn ystod y pythefnos canlynol. Cyfwelwyd ag ugain o gyfranogwyr wyneb yn wyneb a dau dros y ffôn.

Ar ddechrau'r cyfweliad, gofynnwyd i'r cyfranogwyr lunio dwy farn wrth feddwl yn uchel am sut y maent yn dewis yr ymateb buddugol. Mae'r dull meddwl yn uchel yn gofyn i gyfranogwyr leisio eu meddwl wrth ddatrys tasg. Yn yr astudiaeth hon fe'i defnyddiwyd i driongli'r hyn yr oedd cyfranogwyr yn ei gofio am lunio barnau gan arsylwi ar yr agweddau a ystyriwyd ganddynt yn ystod barnau'r cyfweliad. Er enghraifft, cadarnhawyd yr agweddau a grybwyllwyd wrth werthuso sgriptiau yn ystod y rhan meddwl yn uchel gyda'r ymatebion yn ystod y cyfweliad. Ar gyfer yr ymarfer hwn, dewiswyd chwe sgript a oedd wedi'u rhestru'n uwch na'r cyfartaledd yn yr ymarfer BG ac a oedd â gwerthoedd paramedr tebyg. Y bwriad oedd dewis ymatebion na fyddai'n hawdd iawn eu cymharu, gyda'r gobaith o gael mwy o sylwadau gan gyfranogwyr yn ystod y sesiwn meddwl yn uchel.

Ar ôl y sesiwn meddwl yn uchel, gofynnwyd cwestiynau i'r cyfranogwyr am eu profiad beirniadu yn ystod y gweithdy a'u barn am BG. Wrth ystyried pa mor ddefnyddiol oedd BG, gofynnwyd i'r cyfranogwyr ddychmygu system asesu lle byddai athrawon yn beirniadu ymatebion o ysgolion eraill, gan fod hyn yn osgoi rhagfarn sy'n gysylltiedig ag adnabod y dysgwyr. Dywedwyd wrth y cyfranogwyr hefyd i ddychmygu na fyddai'r ymarfer cyfan yn cymryd mwy o amser nag y maent yn ei dreulio'n marcio ymatebion NEA. Rhoddwyd y cyfarwyddyd hwn i osgoi cyfranogwyr yn ymhelaethu ar yr agwedd hon, na fyddai ganddynt ddigon o wybodaeth ar gyfer hyn i lunio barn gywir wrth ystyried costau a buddion y dull.

5. Canlyniadau Barn Gymharol

Mae'r adran hon yn cyflwyno ac yn trafod canlyniadau meintiol y sesiynau beirniadu yn fyr. Ffocws yr astudiaeth oedd archwilio barn athrawon, ond bydd canlyniadau barn gymharol yn helpu i roi'r safbwyntiau hynny yn eu cyd-destun ac yn caniatáu cymariaethau ag astudiaethau tebyg.

5.1 Nifer y barnau

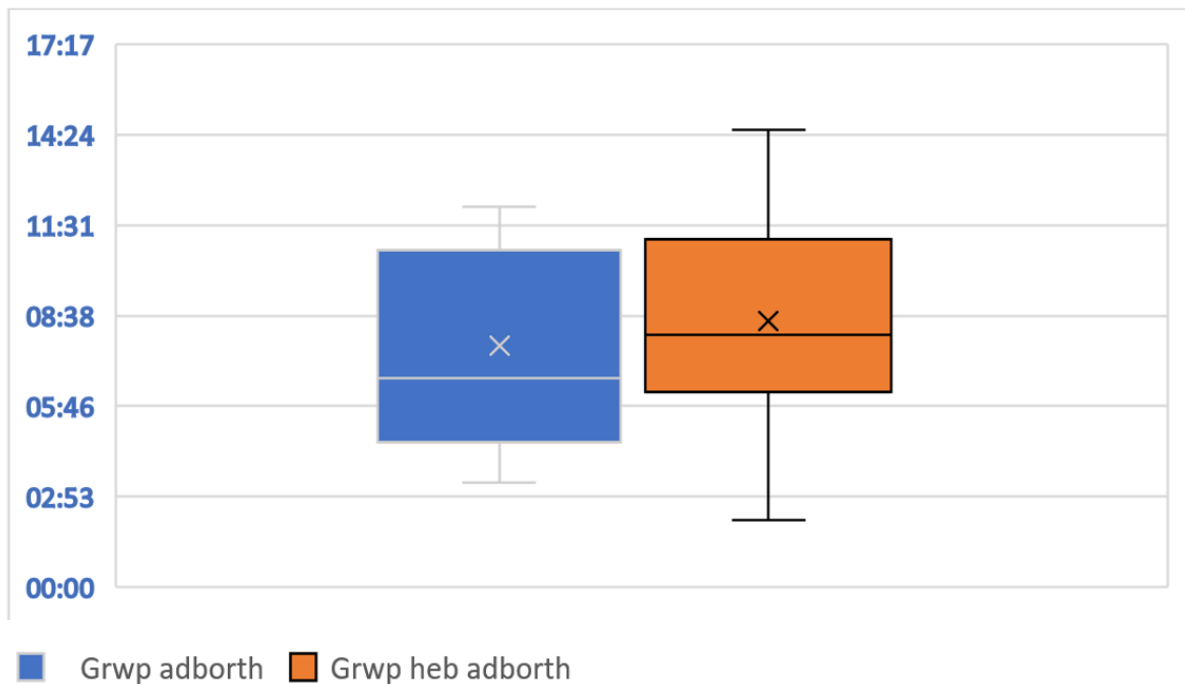
Adeg yr astudiaeth, cyfyngodd meddalwedd RM Compare™ ar nifer y barnau i 21 y beirniad. Gwnaeth y rhan fwyaf o'r cyfranogwyr lunio 21 barn a dod i ben, gan na fyddai'r system yn caniatáu iddynt wneud mwy o farnau; gwnaeth dau gyfranogwr 17 o farnau; gwnaeth un cyfranogwr 20 o farnau; ac roedd un arall rywsut yn gallu osgoi'r cyfyngiad a osodwyd a gwneud 28 o farnau. Gwnaed 230 o farnau ym mhob grŵp (cyfanswm o 460); felly, beirniadwyd pob sgript 10 gwaith ar gyfartaledd. Yr ystod o farnau fesul sgript oedd 17 i 25 yng Ngrŵp NF, a 18 i 24 yng Ngrŵp F.

5.2 Amseroedd beirniadu

Yr amser cymedrig ar gyfer un farn oedd 7 munud 34 eiliad yng Ngrŵp F ac 8 munud 05 eiliad yng Ngrŵp NF (neu 8 munud 52 eiliad ac eithrio amseroedd beirniad a oedd yn beirniadu ar gyfartaledd ymhen 2 funud 8 eiliad³). Dangosir y cymedrau hyn fel croesau yn Ffigur 2 isod. Yr amser beirniadu canolrif ar gyfer Grŵp F oedd 6 munud 34 eiliad ac 8 munud 01 eiliad ar gyfer Grŵp NF. Cyfartaledd cyfartaleddau'r beirniaid oedd 7 munud 35 eiliad ac 8 munud 27 eiliad yn y drefn honno (gan gynnwys yr eithriad 2 funud 8 eiliad). Dengys Ffigur 2 fod cryn orgyffwrdd rhwng yr amser cyfartalog ar gyfer pob beirniad yn y ddau grŵp. Dengys y data fod cyfranogwyr yng Ngrŵp NF yn arafach ar y cyfan a bod ganddynt ystod ehangach o amseroedd beirniadu na'r rhai yng Ngrŵp F, sy'n cadarnhau canfyddiadau goddrychol y rhan fwyaf o gyfranogwyr fod darparu adborth wedi helpu i gyflymu'r broses farnu (gweler adran 6.4.4).

³ Er ei gyflymder, roedd y cyfranogwr a dreuliodd ddwy funud ar gyfartaledd yn beirniadu pob pâr yn gyson â'r beirniaid eraill. Trafodir *infit* a dibynadwyedd ar gyfer yr holl feirniaid a'r sgriptiau yn fanylach yn nes ymlaen.

Ffigur 2: Amser beirniadu cyfartalog ar gyfer pob beirniad (munudau)



Mesurir yr amser ar gyfer un farn yn ôl yr amser a aeth heibio rhwng cynhyrchu'r dyraniad pâr a chofnodi'r penderfyniad. Er mwyn osgoi cofnodi amser segur, fe wnaethom annog cyfranogwyr i allgofnodi o'u cyrifon wrth gymryd seibiant (dim ond ar ôl tua 24 munud o anweithgarwch y byddai'r feddalwedd yn allgofnodi'n awtomatig). Ond byddai rhai cyfranogwyr yn dal i fynd i sgwrs o bryd i'w gilydd (hyd y gwyddom, heb fod yn gysylltiedig â'r farn yr oedd yn rhaid iddynt ei gwneud) gyda chydweithwyr cyfagos heb allgofnodi. Mae hyn yn golygu bod yr amser beirniadu cymedrig yn is na'r hyn a nodwyd uchod, yn ôl pob tebyg.

5.3 Gwerthoedd *misfit* a dibynadwyedd

Mae model ystadegol Bradley-Terry y tu ôl i feddalwedd RM Compare™ yn cyfrifo'r dibynadwyedd a'r gwerthoedd *infit* yn ogystal ag ystadegau eraill sy'n berthnasol i BG. Cyflawnodd y beirniaid yng Ngrŵp F gyfernod dibynadwyedd o 0.88 (+/-0.02). Cyflawnodd y beirniaid yng Ngrŵp F gyfernod dibynadwyedd is o 0.74 (+/-0.02). Er bod y mynegeion dibynadwyedd ar gyfer y ddau grŵp yn uchel, nid oeddent mor uchel ag mewn astudiaethau eraill. Er cymharu, mewn astudiaeth yn beirniadu traethodau daearyddiaeth heb fod yn fwy na dwy dudalen o hyd (o lawysgrifen yn ôl pob tebyg) ac y gellid rhoi uchafswm o 15 marc iddynt, y dibynadwyedd oedd 0.97 ar ôl 12.5 cylch (tua 6.2 barn fesul sgrïpt) a mwy na 0.9 ar ôl dim ond saith cylch (Whitehouse a Pollitt, 2012). Roedd astudiaeth arall yn cymharu ymatebion i un cwestiwn o adran ysgrifennu TGAU Saesneg Haen Uwch. Roedd yr ymatebion yn amrywio o un dudalen i nifer ohonynt, a gellid rhoi uchafswm o 40 marc iddynt. Ar ôl

tua 14 o farnau fesul sgrïpt, cyrhaeddwyd dibynadwyedd o 0.97 (Bramley a Vitello, 2019). Mae'n bosibl bod y dibynadwyedd is yn ein hastudiaeth wedi'i achosi gan fod cyfranogwyr yn gorfod beirniadu ymatebion o ddwy dasg wahanol. Dylid ymchwilio ymhellach i effaith gorfod beirniadu ymatebion ysgrifennu yn seiliedig ar wahanol dasgau ar ddibynadwyedd.

Mae'n bosibl hefyd bod y cyfernodau dibynadwyedd yn ein hastudiaeth yn is am nad ydynt wedi'u chwyddo. Yn dilyn beirniadaeth Bramley (2015) fod ymaddasolrwydd yn chwyddo'r cyfernodau dibynadwyedd, bu Rangel-Smith a Lynch (2018) yn archwilio'r mater ac yn awgrymu ffyrdd o fireinio'r algorithm paru i fynd i'r afael â'r mater. Mae RM wedi gweithredu'r argymhellion ac awgrymir bellach y dylai 15 cylch o farnau ddileu'r chwyddiant dibynadwyedd (Kimbell, 2021). Yn ein hastudiaeth, cynhaliwyd 20 o gylchoedd beirniadu.

Mae'r model ystadegol hefyd yn cyfrifo'r gwahaniaethau rhwng y canlyniadau disgwyliedig a'r canlyniadau a arsylwyd mewn barn benodol ar gyfer pob beirniad a sgrïpt. Gellir safoni'r gwahaniaethau hyn mewn dwy ffordd ac fe'u gelwir yn werthoedd *infit* ac *outfit*, ac maent yn seiliedig ar fodel logistaidd Rasch sy'n cyfateb i'r model a luniwyd gan Thurstone (Pollitt 2012). Defnyddir gwerthoedd *infit* i nodi beirniaid nad oeddent yn gyson â gweddill y grŵp, neu i nodi sgrïptiau a oedd yn gwneud beirniaid yn llai cyson wrth eu beirniadu, sef yr hyn y mae Pollitt (2012) yn ei alw'n *misfitting*. Fel confensiwn, nododd Pollitt (2012) fod gwerthoedd *infit* o fewn dau wyriad safonol o'r cyfartaledd yn cael eu hystyried yn dderbyniol, a bod y rhai sy'n fwy na dau wyriad safonol yn cael eu hystyried yn wahaniaethau neu'n *misfitting* sylweddol. Mae beirniaid y mae eu gwerthoedd *infit* yn mynd y tu hwnt i'r terfynau derbyniol yn gwneud mwy o benderfyniadau annisgwyl na gweddill y grŵp. Gellid dehongli bod ganddynt ddealltwriaeth wahanol o beth yw gwaith o ansawdd da (Bramley, 2007 t. 260). Fodd bynnag, pan wneir dehongliadau o'r fath, dylid cadw mewn cof fod ystadegyn *misfit* yn fesur cymharol gan ei fod yn meintio'r graddau y mae'r penderfyniadau a wneir gan feirniad yn wahanol i'r penderfyniadau a wnaeth beirniaid eraill yn y grŵp. O'r herwydd, nid yw *misfitting* o reidrwydd yn dangos beirniadu gwael (Pollitt, 2012).

Gwerth *infit* cyfartalog y beirniaid yng Ngrŵp F oedd 1.03 gyda gwyriad safonol (GS) 0.345. Roedd gan un cyfranogwr werth *infit* uwch na derbyniol – sef 1.87. Gadawodd y beirniad hwn sylwadau byr a chyffredinol iawn, i'r graddau efallai na fuasent yn ddefnyddiol i helpu gyda'r beirniadu⁴. Mae hyn yn gwrthgyferbynnu â'r hyn a ddangosodd Daal et al. (2016) yn eu hastudiaeth lle'r oedd yr un beirniad *misfitting* (gwerthoedd *infit* uwch) wedi gadael mwy o ddadleuon fesul sylw na beirniaid eraill. Efallai nad hyd y sylwadau mo achos *misfitting*, ond y graddau y gall sylwadau fod yn

⁴ Gwnaeth y cyfranogwyr 204 o sylwadau, hynny yw, gwnaethant 9 sylw ar gyfartaledd fesul ymateb, sy'n golygu bod beirniaid wedi gadael tuag un sylw fesul barn a wnaethant. O gyfanswm y sylwadau, gwnaed 20 yn ogystal â'r rhai cychwynnol. Roedd y sylwadau'n amrywio o ddau air i 64 gair.

ddefnyddiol i lywio barnau presennol neu farnau yn y dyfodol gydag eithafion yn cydberthyn i benderfyniadau mwy annisgwyl. Roedd gwerthoedd *infit* gweddill y beirniaid yn amrywio o 0.57 i 1.14.

Gwerth *infit* cyfartalog y sgriptiau yng Ngrŵp F oedd 0.99 gyda GS o 0.3. Roedd gan bob sgript werth *infit* derbyniol yn amrywio o 0.59 i 1.54, ac eithrio un sgript ag *infit* isel o 0.08 ac a ddosbarthwyd yn olaf, h.y. 23/23 (wedi'i dosbarthu 6/23 yng Ngrŵp NF). Rhoddwyd 12 marc i'r sgript hon. Sylwch mai hon yw'r unig sgript na wnaeth y beirniaid unrhyw sylw amdani. Ar ôl cael caniatâd yr athro a ddarparodd y sgript hon, fe'i rhannwyd gyda chyfranogwyr ynghyd â dwy sgript arall a oedd hefyd ymhellach i ffwrdd yn y drefn a ddeilliodd o hynny o gymharu â graddio marciau. Dywedasom wrth yr athrawon fod gan y tair sgript rai gwahaniaethau mwy amlwg o ran graddio a gofynnwyd iddynt ymchwilio i weld a oedd unrhyw beth a fyddai'n gwneud iddynt asesu'r sgriptiau hyn yn wahanol yn ystod BG nag wrth farcio. Atebodd tri chyfranogwr, ac nid oedd lefel y manylder yn ddigon i gael unrhyw fewnwelediad i'r esboniad posibl am y *misfitting*. Gwerth *infit* cyfartalog y beirniaid yng Ngrŵp NF oedd 1 (GS o 0.22). Roedd gan un beirniad werth *misfit* uwch na derbyniol o 1.47. Roedd gwerthoedd *infit* gweddill y beirniaid yn amrywio o 0.76 i 1.42. Roedd gan y beirniad gyda'r amser beirniadu cyfartalog o 2 funud ac 08 eiliad werth *infit* derbyniol o 0.92.

Gwerth *infit* cyfartalog y sgriptiau yng Ngrŵp F oedd 0.96 gyda GS o 0.265. Roedd gan bob sgript werthoedd *infit* derbyniol yn amrywio o 0.64 i 1.38, ac eithrio un sgript a chanddi werth *infit* isel o 0.08 ac a ddosbarthwyd yn olaf (wedi'i dosbarthu 15/23 yng Ngrŵp F). Dyfarnwyd 12 marc i'r sgript hon ac mae'n dod o ysgol wahanol i'r sgript gyda'r gwerth *infit* isel yng Ngrŵp F. Dywedodd y beirniaid yng Ngrŵp F fod y sgript hon yn dangos gwybodaeth dda a disgrifiad cyd-destunol ond ei bod yn wan o ran gwerthuso a dadansoddi. Gallai'r anghysondeb hwn esbonio'n rhannol y gwahaniaeth barn ar y sgript hon.

Roedd gan y ddau feirniaid yr oedd eu gwerthoedd *infit* yn annerbyniol o uchel fwy na saith mlynedd o brofiad marcio. Nid yw eu gwerthoedd *misfit* o reidrwydd yn dangos iddynt wneud gwaith gwaeth na'r lleill, ond bod ganddynt wahanol farn am ansawdd yr ymatebion. Un darn diddorol o ddata ynghylch y beirniaid *misfitting* yw y byddai cyfranogwyr yn yr astudiaeth hon fel arfer yn llwyddo i lunio dwy i bedair barn yn ystod y pedwar cylch cyntaf, hynny yw cyn y byddai'r ymaddasolrwydd a'r cadwyno'n dechrau. Cynhaliodd y beirniad gyda'r gwerth *misfitting* annerbyniol yng Ngrŵp F 11 o farnau allan o 21 yn ystod y pedwar cylch cyntaf. Ar y llaw arall, cynhaliodd y beirniad â'r gwerth *misfit* uchaf yng Ngrŵp NF 5 allan o 20 o farnau yn y pedwar cylch cyntaf. At hynny, cynhaliodd beirniad o'r un grŵp 11 allan o 20 barn yn y pedwar cylch cyntaf (yn debyg i'r beirniad *misfitting* uchaf yng Ngrŵp F) ond roedd ganddo werth *infit* derbyniol o 0.9. O ystyried hyn, mae gan nifer y barnau a lunnir cyn ymaddasolrwydd a chadwyno berthynas gymhleth â *misfit*.

Er bod rhai astudiaethau BG (e.e. Newhouse, 2014; McMahon a Jones, 2015) yn defnyddio'r rheol dau wyriad safonol a awgrymwyd gan Pollitt, yn nadansoddiad Rasch defnyddir rheolau eraill hefyd i nodi gwerthoedd *misfitting*. Er enghraifft, un dull arall yw ystyried y gwerthoedd sydd y tu allan i ystod 0.70-1.30 fel rhai *misfitting* (gweler er enghraifft, Hodge a Morgan, 2017). Os defnyddir y rheol hon, yna yng ngrŵp NF byddai dau feirniad â gwerthoedd *infit* yn fwy nag 1.30 a dwy sgript islaw'r gwerth toriad o 0.7. Yng Ngrŵp F byddai un beirniad gyda'r *infit* o 0.64 a dau feirniad gyda'r *infit* yn fwy nag 1.3. Byddai tair sgript hefyd gyda gwerthoedd *infit* islaw, a thair sgript gyda gwerthoedd *infit* uwchlaw'r pwyntiau torri i ffwrdd.

Nid oedd cydberthynas rhwng graddio darllenadwyedd canfyddedig yr ymatebion ysgrifenedig a'u safle BG, sy'n awgrymu nad oedd darllenadwyaeth y llawysgrifen yn effeithio ar ddilysrwydd y cymariaethau sgript.

Y gydberthynas rhwng y drefn restrol yn seiliedig ar farciau a ddyfarnwyd i'r sgriptiau a'r drefn restrol yn seiliedig ar BG ar gyfer Grŵp F oedd 0.5 ($p=0.014$) a 0.73 ($p<0.01$) ar gyfer Grŵp NF. Y prawf cydberthyniad a berfformiwyd oedd cydberthyniad rhestrol Spearman dwy-gynffon. Er mwyn eithrio canlyniadau cyfartal ar gyfer y drefn sy'n seiliedig ar farciau, ym mhob grŵp marcio fe ddsbarthwyd yr ymateb uchaf gyda'r un nifer o farciau a ddsbarthwyd uchaf yn y BG, ac yn y blaen. Pan gaiff canlyniadau cyfartal eu cynnwys, mae'r cyfernodau cydberthyniad yn llawer is ac nid ydynt yn ystadegol wahanol i sero: ar gyfer Grŵp F y cyfernod cydberthyniad oedd 0.25 ($p=0.24$) ac ar gyfer Grŵp NF roedd yn 0.31 ($p=0.15$). Fe wnaethom gyfrifiannu lefel hyder 95% cyfernodau cydberthyniad rhestrog Spearman drwy ymlwytho gan ddefnyddio pecyn RVAideMemoire yn R. Ar gyfer Grŵp F roedd y CI rhwng -0.18ai 0.64 ac ar gyfer Grŵp NF roedd y CI rhwng -0.18 a 0.71.

Yn astudiaeth Whitehouse a Pollitt (2012) a grybwyllir uchod, y gydberthynas a gyflawnwyd gyda marcio traddodiadol oedd 0.63 ar ôl tua chwe barn fesul sgript. Yn astudiaeth Bramley a Vitello (2019) roedd cydberthynas Pearson o ganlyniadau BG â marciau'r prif arholwyr yn 0.74 (Bramley a Vitello, 2019). Mewn astudiaeth a oedd yn cynnwys atebion byrrach i arholiadau Hanes safon UG, roedd y gydberthynas yn amrywio o 0.84 i 0.86 (Holmes et al., 2018, t.22).

Yn yr astudiaeth hon, nid yw canlyniadau'r profion cydberthyniad yn argyhoeddiadol. Ar y naill law, mae'r cyfernodau cydberthyniad pan gaiff canlyniadau cyfartal eu cynnwys yn awgrymu perthynas wan rhwng y drefn restrol o BG a marcio. Ar y llaw arall, mae ansicrwydd mawr ynghylch y cyfernodau hyn sy'n dangos y gallai cydberthynas fod mor uchel â 0.71 ac mor isel â -0.18. Mae'r ansicrwydd hwn yn debygol o gael ei achosi gan y meintiau sampl bach, clystyru'r sgriptiau mewn ystod fach o farciau a'r nifer fawr o safleoedd cyfartal yn y raddfa farciau.

Mae'n bosibl mai un ffactor sy'n effeithio ar y dilysrwydd yw hyd y sgrïptiau. Mae'n bosibl mai po hiraf yw'r sgrïptiau, y mwyaf anodd yw sicrhau cydberthynas uwch, gan dybio y dylid gwerthuso a chadw mwy o wybodaeth mewn cof wrth gymharu'r sgrïptiau'n gyfannol, yn wahanol i farcio dadansoddol lle rhoddir marciau ar gyfer gwahanol agweddau cyn eu cyfuno'n farc terfynol. Gallai gorfod beirniadu sgrïptiau ar wahanol bynciau fod wedi effeithio ar ddilysrwydd hefyd. Mae'n bosibl hefyd, yn wahanol i anodi, sy'n targedu amcanion asesu yn benodol, bod gwneud sylwadau agored, byr heb arweiniad wedi gwneud i'r sylwadau hynny wyro oddi wrth y meini prawf marcio gwirioneddol sydd wedyn yn cael ei ddwysáu pan fydd nifer y barnau'n cynyddu. Felly, mae angen ymchwilio ymhellach i ddilysrwydd beirniadu ymatebion ysgrifennu estynedig.

Roedd y cyfernodau cydberthyniad (wrth eithrio a chynnwys canlyniadau cyfartal) yn is yng Ngrŵp F. Dywedodd beirniaid yn y ddau grŵp eu bod wedi defnyddio'r cynllun marcio a'r amcanion asesu i lywio eu penderfyniadau. Wrth i ni drafod ymhellach, yr unig eithriad i hyn allasai fod y cyfranogwyr nad oeddent byth yn marcio NEA TGAU nac arholiadau. Pan rannwyd y cyfranogwyr yn grwpiau, ceisiom sicrhau eu bod yn fras gynrychioli'r math o ysgol a blynyddoedd o brofiad. Ar hap, dyrannwyd yr holl athrawon nad oeddent erioed wedi marcio TGAU i Grŵp F. Gallai hyn fod yn esboniad credadwy ynghylch pam yr oedd dilysrwydd cydgyfeiriol yn is i'r grŵp hwn. Un esboniad arall am yr anghysondeb hwn efallai yw'r ffaith bod gwneud sylwadau'n effeithio ar ddilysrwydd barnau. Weithiau defnyddiwyd sylwadau a wnaed yn gynharach wrth wneud penderfyniadau, a gallai hyn fod yn esboniad arall.

6. Profiad cyfranogwyr o BG a'r agweddau a ystyriwyd ganddynt wrth feirniadu

Yn gyffredinol, soniodd bron pob un o'r cyfranogwyr eu bod wedi mwynhau'r ymarfer. Roedd y cyfranogwyr yn cyferbynnu BG â marcio confensiynol ac yn credu bod BG yn brofiad mwy diddorol. Fel y byddwn yn ei ddangos yn yr adran nesaf, roeddent yn meddwl bod marcio'n anoddach yn wybyddol. Yn ogystal, mae'r dull mwy strwythuredig yn gwneud i farcio deimlo ar adegau yn 'ddiflas', yn 'anniddorol', 'megis ticio bocsys'. I'r gwrthwyneb, ystyriwyd bod BG yn fwy diddorol oherwydd y dull cyfannol o asesu. At hynny, credai cyfranogwyr y byddai BG yn annog dysgwyr i ysgrifennu ymatebion mwy creadigol a llai fformwlaig, a fyddai'n gwneud asesu ymatebion hyd yn oed yn fwy diddorol, er nad ydym wedi ystyried a fyddai hyn yn gwneud beirniadu'n fwy heriol.

Mae'r math mwy traddodiadol o farcio yn teimlo hyd yn oed yn fwy difflach nag o'r blaen, rywsut.

Roeddwn i'n disgwyl i mi fy hun ddiflasu braidd. Clicio, clicio, clicio. Ond mewn gwirionedd,... ar ôl i mi ddod i'r arfer ag ef, roeddwn i'n iawn.

Roedd cyfranogwyr o'r farn eu bod wedi dod i arfer â'r dull asesu newydd yn gyflym a'u bod wedi gwella arno wrth iddynt wneud mwy o farnau. Roeddent yn teimlo bod cywirdeb eu penderfyniadau wedi aros yr un fath drwy gydol y dydd. Fodd bynnag, soniodd rhai cyfranogwyr y gallai'r sŵn cefndir yn yr ystafell fod wedi effeithio ar eu canolbwyntio.

Teimlai'r cyfranogwyr ei bod yn anos beirniadu erbyn diwedd y gweithdy. Roeddent o'r farn bod hyn yn 'arafu naturiol' a fyddai'n digwydd yn ystod marcio hefyd, ac nad oedd yn waeth nag wrth farcio. Yn hyn o beth, soniodd cyfranogwyr fod angen sicrhau bod beirniaid yn cymryd seibiannau, a ystyriwyd hyd yn oed yn bwysicach o ystyried bod ymatebion ar y sgrin.

6.1. Dulliau o ddarllen ymatebion

Yn seiliedig ar arsylwadau yn ystod y sesiynau meddwl yn uchel ac ar atgofion cyfranogwyr yn ystod cyfweiliadau, canfuwyd bod beirniaid yn defnyddio sawl dull o ymgyfarwyddo â'r ddau ymateb cyn llunio eu barn. Darllenodd rhai cyfranogwyr un ymateb ac yna darllen yr ail un. Byddai grŵp arall o gyfranogwyr yn darllen ychydig o bob ymateb yn olynol. Byddai'r rhan fwyaf o'r beirniaid hyn fel arfer yn darllen dadansoddiad a gwerthusiad un neu ddwy ffynhonnell o bob ymateb, er y byddai rhai yn darllen un dudalen. Defnyddiwyd y dull olaf hwn fel ffordd o ddod o hyd yn gyflym i'r man lle'r oeddent wedi dod i ben yn yr ymateb blaenorol. Byddai trydydd grŵp o feirniaid yn 'cael ychydig o deimlad' yn gyntaf o un ymateb o leiaf cyn darllen yr ymateb(ion) yn llawn. Er enghraifft, byddent yn darllen y cyflwyniad neu'r dudalen gyntaf, yn mynd i'r casgliad ac yna'n darllen yr ymateb cyfan.

Wrth ddarllen, byddai'r cyfranogwyr weithiau'n defnyddio techneg bwrw golwg cyflym. Yn yr achos hwn, dim ond rhai elfennau o bob ymateb y byddent yn eu dewis, cyn penderfynu pa un oedd yn well. Y rhan fwyaf o'r amser, roedd bwrw golwg cyflym yn golygu darllen y cyflwyniad, y casgliad ac un neu ddwy ffynhonnell yn fwy sylwgar, a dim ond bwrw golwg cyflym ar weddill yr ymateb. Mae'n debyg bod y dull hwn o fwrw golwg cyflym yn tybio ymateb wedi'i strwythuro'n dda ac efallai na fydd mor ddefnyddiol ar gyfer ymatebion llai strwythuredig.

Un dull arall o fwrw golwg cyflym oedd darllen yr ymateb cyfan yn frysio wrth ddewis rhywfaint o gynnwys ar gyfer darllen manylach. Dywedodd cyfranogwyr eu bod yn defnyddio'r dull olaf hwn wrth farcio, er eu bod yn teimlo bod angen iddynt fod yn fwy trylwyr wrth farcio oherwydd bod angen iddynt anodi. Mae'n ymddangos bod y math hwn o fwrw golwg cyflym yn dibynnu ar adnabod y geiriau 'allweddol' neu 'gyffredin' ac yna chwyddo i mewn i'r testun o amgylch y geiriau hynny.

Wrth esbonio eu dull o fwrw golwg cyflym, ni allai llawer o'r cyfranogwyr ei ddisgrifio gan ei fod yn rhywbeth y byddent yn ei wneud yn reddfól.

Nid oes gennym ddata manwl i ymchwilio i weld a gafodd gwahanol ddulliau o ddarllen yr ymatebion unrhyw effaith ar ansawdd barnau. Gellid ymchwilio ymhellach i'r ffordd y mae beirniaid yn darllen yr ymatebion er mwyn canfod a oes un ffordd fwyaf effeithlon o ddarllen yr ymatebion, waeth beth fo'u dewisiadau personol.

6.2. Yr angen i ddarllen ymatebion yn llawn

Un penderfyniad yr oedd yn rhaid i'r cyfranogwyr ei wneud oedd a allent lunio'r farn ar ôl darllen neu fwrw golwg cyflym ar ran o ymateb yn unig.

Awgrymodd cyfran fach o'r cyfranogwyr fod angen i feirniad ddarllen y ddau ymateb yn llawn er mwyn gwneud penderfyniad teg. Dywedodd y cyfranogwyr hyn y byddent yn parhau i ddarllen neu, o leiaf, bwrw golwg cyflym ar yr ymatebion yn llawn, hyd yn oed pe bai'n amlwg bod un ymateb yn well na'r llall. Roeddent yn teimlo y gall ansawdd ymateb newid, felly ni fyddai'n deg llunio barn ar ôl ymgyfarwyddo â rhan ohono'n unig.

Fedrwn i byth, byth ddarllen ychydig bach yn unig a rhoi marc achos 'dw i ddim yn meddwl bod hynny'n deg.

Roedd rhywfaint o dystiolaeth y byddai o leiaf rai o'r cyfranogwyr hyn yn darllen yr ymateb cyfan yn ôl eu harfer, ond gyda mwy o brofiad beirniadu gallent newid eu dull gweithredu.

P: Rhaid imi ddarllen y cyfan. Ond rwyf 99.9% yn siŵr y bydd B yn well.

I. Felly, pam ydych chi'n meddwl bod angen i chi ddarllen y cyfan?

P: Oherwydd bod angen i chi wneud yn siŵr.

I ... os ydw i'n onest beth fydda i'n ei wneud nawr yw sut y byddwn i'n marcio.

Ond roedd mwyafrif llethol y cyfranogwyr (gan gynnwys y ddau â *misfit* annerbyniol) yn cytuno nad oedd angen darllen na bwrw golwg cyflym ar ymateb yn ei gyfanrwydd pan oedd yn amlwg pa un oedd yn well.

Mewn achosion o'r fath, byddai'r cyfranogwyr hyn naill ai'n rhoi'r gorau i ddarllen yr ymateb neu'n darllen rhai elfennau ychwanegol yn ddetholus (e.e. casgliad neu ffynhonnell arall). Mae'n debyg bod y penderfyniad i gael y gwiriad ychwanegol hwn yn dibynnu nid yn unig ar gryfder eu hargyhoeddiad, ond hefyd ar ddewis pob beirniad.

Byddwn ar hyn o bryd yn mynd i'r farn ar y diwedd dim ond i weld ei ansawdd o'r blaen, a byddaf yn gweld a oes angen mynd yn ôl at weddill yr ateb.

Dywedodd cyfranogwyr a oedd o'r farn nad oedd bob amser angen darllen pob ymateb yn llwyr nad ydynt yn mabwysiadu'r un dull wrth farcio oherwydd eu bod yn teimlo bod pennu marc yn wahanol i benderfynu pa ymateb oedd yn well.

Gellir ffurfio'r argyhoeddiad fod un ymateb yn well na'r llall yn gyflym. Er enghraifft, yn ystod sesiynau meddwl yn uchel, roedd cyfranogwyr eisoes yn gallu gwneud penderfyniad ar ôl darllen y dadansoddiad o un neu ddwy ffynhonnell yn yr ail ymateb yn unig.

Ydy, mae wedi ysgrifennu wyth ochr, ond hyd yn oed wedi darllen y dudalen gyntaf a hanner, gallwch ddweud bod hwn yn well oherwydd ei fod yn rhyngweithio â'r ffynhonnell.

Wedi darllen dwy dudalen, mae hon yn sylweddol well na hon. Does dim angen i mi ddarllen y gweddill hyd yn oed, oherwydd gallaf weld hynny. Ac erbyn diwedd pob paragraff gallwch weld eu bod yn gwneud y meddwl uwch hwnnw o ran y dadansoddiad.

Rydych chi'n cael syniad o'r ddwy ffynhonnell gyntaf. Byddwn i'n dweud, 95% neu 99% o'r amser, y bydd ffynhonnell tri a phedwar yr un fath â ffynhonnell un a dau.

Y cwestiwn pwysig yw a ydy darllen yr ymatebion yn anghyflawn yn effeithio ar ansawdd y barnau. Roedd bron pob un o'r cyfranogwyr yn hyderus ynghylch ansawdd eu penderfyniadau. Roedd hyn yn bennaf am eu bod o'r farn bod eu profiad marcio wedi eu helpu ac am eu bod yn teimlo bod cymharu yn haws na marcio, sy'n gofyn 'pennu marciau gwirioneddol'. Mae'r ystadegau *misfit* dau wyriad safonol a gyflwynwyd yn gynharach a'r gydbertynas rhwng canlyniadau BG a'r marciau i'w gweld yn dangos nad oedd y darlleniad anghyflawn yn cael effaith negyddol yn arbennig. Fodd bynnag, byddai angen ymchwilio mwy i'r ffenomenon hon, yn enwedig o ystyried bod mwy o *misfitting* wedi'i nodi wrth gymhwyso synnwyr y fawd 0.7-1.3.

Roedd beirniadu ymatebion wedi'u cadwyno yn achos penodol o ddarllen anghyflawn. Dywedodd yr holl gyfranogwyr na fyddent yn darllen nac yn edrych ar yr ymateb a ddangoswyd yr ail dro yn olynol oni bai bod yr ymatebion o ansawdd tebyg, gan ei gwneud yn anos gwneud penderfyniad terfynol. Yn yr achos hwn, byddai cyfranogwyr yn dychwelyd i'r ymateb wedi'i gadwyno i wirio rhannau ohono'n ddetholus.

Roedd y cyfranogwyr yn hoffi cadwyno gan eu bod yn credu ei fod yn eu helpu i feirniadu'n gyflymach. Roeddent hefyd yn teimlo nad oedd cadwyno'n gwneud eu barn yn rhagfarnllyd.

Byddai'r feddalwedd BG bob amser yn rhoi'r ymateb wedi'i gadwyno yn gyntaf. Ar gyfer pob grŵp, cyflawnwyd prawf ystadegol⁵ i bennu cryfder y dystiolaeth ystadegol bod rhagfarn wrth ddewis yr ymateb cyntaf dros yr ail ymateb.

Dilëwyd y penderfyniadau o'r pum cylch cyntaf: yn y pedwar cylch cyntaf nid yw'r cadwyno'n bodoli ac yn y pumed nid yw'n amlwg eto. Yng ngweddiill y rowndiau, dewiswyd yr ymateb cyntaf (wedi'i gadwyno) yng Ngrŵp F 84 allan o 173 o weithiau neu mewn cyfrannedd o 0.49 (99% CI: 0.388-0.584; P – 0.704) ac yng Ngrŵp NF 97 allan o 173 gwaith neu mewn cyfrannedd o 0.56 (99% CI: 0.463- 0.658; P – 0.110). Nid oedd y gwerth P yn ystadegol arwyddocaol yn y naill achos na'r llall ar y lefel arwyddocâd o 1%. Felly, nid oedd unrhyw dystiolaeth ystadegol gref yn yr astudiaeth hon fod y cadwyno wedi arwain at ragfarn wrth ddewis y gwrthrych wedi'i gadwyno yn amlach (er y gallai maint y sampl gyfyngu ar y dadansoddiad).

Soniodd Pollitt a Crisp (2004), a oedd hefyd yn gweithredu cadwyno, nad oedd unrhyw ragfarn wedi'i chanfod mewn unrhyw astudiaethau (gan gyfeirio'n fwyaf tebygol at faes addysgol) wrth weithredu'r dull hwn. Fodd bynnag, dangosodd astudiaeth a oedd yn defnyddio cymhariaeth pâr o wrthrychau sy'n ymwneud â nwyddau economaidd ac amgylcheddol, pan wynebri y sgriptiau fwy o weithiau (gan gynnig y posibilrwydd o'u cofio) fod rhagdybiaeth annibynnol BG yn agored i'w gwestiynu (Brown a Peterson, 2009, t. 82). Mae angen ymchwilio ymhellach i'r mater hwn.

6.3. Agweddau a ystyriwyd wrth wneud y penderfyniadau

Dywedodd cyfranogwyr a oedd yn gyfarwydd â'r cynllun marcio eu bod yn ei ddefnyddio i benderfynu pa ymateb oedd yn well. Byddai'r cyfranogwyr hyn yn ystyried yr un agweddau ag y byddent wrth farcio. Er enghraifft, byddai cyfranogwyr yn gwirio a oedd dysgwyr yn ateb cwestiynau defnyddioldeb a dibynadwyedd y ffynonellau a archwiliwyd; byddent yn asesu dealltwriaeth dysgwyr o'r cyd-destun; ac arddull eu hesboniad. Fodd bynnag, gan eu bod yn adnabod y cynllun marcio'n dda, nid oedd yn rhaid iddynt edrych arno wrth feirniadu.

Rydych chi'n dal i wneud yr un peth, rydych chi'n dal i'w ddarllen ac rydych chi'n dal i gymhwyso'r cynllun marcio, felly'r un broses yw hi o hyd.

⁵ Prawf o gyfran finomaidd gan dybio brasamcan arferol y bodlonwyd y tybiaethau ($np > 5$ ac $nq > 5$) ar ei gyfer.

Doeddwn i ddim yn meddwl bod angen y cynllun marcio arnom oherwydd roeddem i gyd yn gwybod beth oedd yn cael ei gynnwys yn y traethawd.

Nid oedd tri o'r cyfranogwyr o leiaf allan o'r pump nad oedd ganddynt unrhyw brofiad o farcio ymatebion TGAU yn defnyddio'r cynllun marcio a beirniadwyd yr ymatebion ganddynt mewn perthynas â'u dealltwriaeth o ymateb ysgrifenedig da. Roedd gwerthoedd *infit* y pum cyfranogwr hyn o fewn yr ystod dderbyniol.

Roedd [y cydweithiwr a eisteddodd yn agos ataf] a minnau'n dod o golegau lle nad oes gennym y cynllun marcio hwnnw; doedden ni ddim yn adnabod y cynllun marcio hwnnw. Cawsom olwg gyflym drosto cyn i ni ddechrau [beirniadu], ond aethom gyda'n greddf a'r hyn yr oeddem yn ei ystyried yn hanesydd da, sef rhywun a all gefnogi deunydd ffynhonnell gyda'i wybodaeth.

Mae Tabl 3 isod yn cyflwyno'r agweddau a gynhwyswyd yn y sylwadau a wnaed gan gyfranogwyr yng Ngrŵp F gyda data ychwanegol ar gyfer beirniaid nad oeddent erioed wedi marcio TGAU. Soniodd cyfranogwyr o'r ddau grŵp hefyd am yr agweddau hyn yn ystod cyfweiliadau a rhoddir ystyriaeth bellach i rai ohonynt yn yr is-adrannau dilynol.

Tabl 3: Agweddau a grybwyllwyd mewn sylwadau gan Grŵp F

Agwedd	Cyfanswm y sylwadau	Canran o'r holl sylwadau y sonnir am yr agwedd ynddynt	Nifer y sylwadau a wnaed gan feirniaid nad oeddent yn marcio TGAU (NGJ)	Canran y sylwadau y sonnir am yr agwedd ynddynt gan NGJ
Gwerthuso	82	40	28	34
Gwybodaeth/gwybodaeth gyd-destunol	79	39	41	49
Dadansoddi	53	26	30	36
Casgliad	52	26	18	22
Dibynadwyedd/defnyddioldeb	37	18	27	33
Cynnwys ffynhonnell	21	10	6	7
Ateb (heb fod yn) mecanistig/fformiwläig	18	9	10	12
Cyflwyniad	9	4	2	2
Strwythur	5	3	3	4
Cyfeiriad at Amcanion Asesu (AA)	2	1	-	-
Sillafu	2	1	1	1
Sylwadau cyffredinol (e.e. ymdrechion i ateb y cwestiwn; dilys, ond generig; heb ganolbwyntio 100% ar y cwestiwn a osodwyd)	40	20	15	18

Caiff yr holl agweddau a grybwyllwyd yn y sylwadau, ac eithrio'r rhai yn cyfeirio at yr ysgrifennu mecanistig, y cyflwyniad a'r strwythur, eu hadlewyrchu yn y cynllun marcio. Cyfeiriodd pob beirniad (ac eithrio un a adawodd sylwadau un neu ddau air) at wahanol agweddau yn eu sylwadau ac nid oedd yn ymddangos eu bod yn canolbwyntio ar rai agweddau yn unig. Mae'n debygol eu bod wedi sôn am yr agweddau a oedd fwyaf amlwg yn y sgrïpt y gwnaethant sylwadau arni.

Fel y gwelir yn Nhabl 3, roedd beirniaid nad oeddent erioed wedi marcio TGAU, ac a oedd ond wedi marcio papurau Safon Uwch, yn cyd-fynd yn gyffredinol â'r grŵp cyfan ynglŷn â'r agweddau a grybwyllwyd ganddynt yn eu sylwadau. Gellir disgwyl hyn o ystyried bod cynllun marcio NEA Safon Uwch Hanes yn pwysleisio'r meini prawf dadansoddi a gwerthuso hefyd. Mae'r iaith a ddefnyddiwyd yn y sylwadau yn

awgrymu bod y beirniaid hyn, i ryw raddau, yn cael eu harwain gan y cynllun marcio hwn. Er enghraifft, soniodd y beirniaid hyn a oedd y sgriptiau'n llunio neu'n cynnwys barn a gefnogwyd, sef un agwedd a bwysleisir yn y fanyleb Safon Uwch. O gymharu, dim ond dau o'r chwe beirniad arall a soniodd am farn yn eu sylwadau. Rhoddodd y pum beirniad hyn fwy o sylw hefyd i'r agwedd wybodaeth sydd eto i'w gweld yn cael ei chrybwyll a'i phwysoli'n fwy yn y fanyleb Safon Uwch. Gallai'r arsylwadau hyn esbonio dilysrwydd cydgyfeiriol isel (pan eithriwyd canlyniadau cyfartal) ar gyfer y grŵp hwn, a chadarnhau hefyd y byddai marcwyr yn dibynnu ar gynlluniau marcio wrth feirniadu.

Nid yw'n glir pa bwysiad a roddwyd ar yr agweddau a grybwyllwyd yn y sylwadau. Fodd bynnag, mae'n rhesymol tybio eu bod yn chwarae o leiaf rhywfaint o rôl. O'r tabl uchod a'r sylwadau a wnaed yn ystod y cyfweiliadau, mae'n amlwg bod cyfranogwyr yn dilyn yr agweddau a gynhwyswyd yn y cynllun marcio, sy'n awgrymu y gallai BG fod yn ddull dilys o asesu ymatebion ysgrifennu estynedig, ond dylid cadarnhau hyn yn feintiol drwy ymchwil gyda chynllun mwy cadarn.

Hyd yn oed os defnyddiodd y cyfranogwyr y cynllun marcio a'r amcanion asesu fel pwyntiau cyfeirio, roeddent yn teimlo bod y broses o wneud penderfyniadau yn ystod BG yn wahanol i farcio. Wrth gymharu, cymerodd cyfranogwyr ddull mwy cyfannol o asesu'r ymatebion am na fyddai'n effeithlon nac yn rhesymol cadw'r dull wedi'i atomeiddio y byddent yn ei ddefnyddio wrth farcio.

Nid oedd cyfranogwyr o'r farn bod defnyddio'r dull cyfannol hwn yn effeithio'n negyddol ar eu penderfyniadau. Soniwyd hefyd na ddylai asesu cyfannol fod yn anodd ei fabwysiadu, oherwydd, i ryw raddau, fe'i defnyddir eisoes wrth farcio rhai pynciau. Rhoddwyd asesiadau Safon Uwch Hanes a Bagloriaeth Cymru fel enghreifftiau o ddull gweithredu o'r fath. Fodd bynnag, i rai, roedd y ffordd newydd hon o edrych ar yr ymatebion yn teimlo'n 'hynod anodd' ar y dechrau, ond roeddent yn gallu addasu'n gyflym.

Er mwyn dod i'w penderfyniadau, defnyddiodd cyfranogwyr ddulliau gwahanol. Er enghraifft, atebodd rhai cyfranogwyr y cwestiwn, 'Beth ydw i'n ei feddwl am y traethawd hwn?'. Yna defnyddiwyd yr argraff gyffredinol hon i benderfynu pa ymateb oedd yn well. Yn debyg i'r canfyddiadau ym Marshal et al. (2020), byddai cyfranogwyr eraill yn neilltuo band i ymatebion neu, yn achos dau gyfranogwr, brasamcan o farc.

Beth bynnag fo'u dull gweithredu, mae'n ymddangos nad oedd llawer o gyfranogwyr yn ymwybodol o'r holl agweddau a effeithiodd ar eu penderfyniadau. Soniodd cyfranogwyr yn aml yn hyn o beth am eu dibyniaeth ar deimladau greddf. Roedden nhw'n ei chael hi'n anodd disgrifio sut roedd greddf yn gweithio neu beth oedd hi. Wrth ei disgrifio byddent yn defnyddio geiriau fel 'canolbwyntio'n llwyr' a 'chael teimlad am y gwaith'.

Roedd y cyfranogwyr yn siŵr bod y reddf hon am ymateb yn dod gyda phrofiad o asesu ymatebion. Felly, roeddent o'r farn y gallai athrawon newydd gymhwyso gael mwy o anawsterau o ran bod yn gyson os nad ydynt yn cael eu harwain gan y cynllun marcio.

6.3.1. Gwerthuso a dadansoddi dibynadwyedd a defnyddioldeb y ffynonellau

Yn ystod y cyfweiliadau, soniodd cyfranogwyr am ddadansoddi a gwerthuso dibynadwyedd a defnyddioldeb y ffynonellau hanesyddol fel yr agweddau pwysicaf wrth feirniadu'r ymatebion. Soniwyd am ddibynadwyedd a defnyddioldeb mewn 37% o'r sylwadau, tra soniwyd am werthuso a dadansoddi mewn tua dwy ran o dair o'r sylwadau (40% a 26% yn y drefn honno). Mae hyn yn adlewyrchu'n fras y cynllun marcio lle sonnir am ddadansoddi a gwerthuso fel rhan o AA3, gyda dadansoddiad hefyd yn cael ei grybwyll o dan AA2.

Wrth ystyried pa mor dda yr oedd dysgwyr yn gwerthuso ac yn dadansoddi defnyddioldeb a dibynadwyedd y ffynonellau, dywedodd cyfranogwyr eu bod yn defnyddio'r un dull y byddent yn ei ddilyn wrth farcio. Er enghraifft, wrth ystyried pa mor dda y dadansoddodd y dysgwr ddibynadwyedd ffynhonnell, byddai'r cyfranogwyr yn asesu a oedd yn ystyried rhagfarn, elfennau o orliwio, a ffactorau eraill a allai effeithio ar ddibynadwyedd (e.e. propaganda neu ddiben y ffynhonnell).

Byddai swmp y marc wedyn yn edrych ar y tarddiad, gan weld pwy a'i hysgrifennodd a pham, a sut y daeth i'r amlwg bod eu casgliadau'n dod allan fel y maent. Y dyddiad - mae hwnnw yn y tarddiad hefyd. A yw'r dyddiad yn arwyddocaol i'r rheswm pam y gwnaeth ei ysgrifennu? Felly, byddai hynny i gyd yn dod i mewn yno ac yna mae gennych y diben wedyn. Pwy yw'r gynulleidfa? Pam maen nhw'n ei ysgrifennu? Beth maen nhw'n ceisio'i gyfleu?

6.3.2. Gwybodaeth

Dim ond dau farc am wybodaeth y gellir eu dyfarnu i ymatebion rhan (a). Mae'r ddau farc hyn hefyd yn cynnwys agweddau ar ansawdd y cyfathrebu ysgrifenedig. Teimlai'r cyfranogwyr, yn ystod marcio, fod y ddau farc hyn yn cael eu rhoi'n hawdd fel arfer ac, er mwyn eu hennill, ni ddisgwylir i'r dysgwr fod â lefel fawr o wybodaeth.

Felly, pan fyddaf yn darllen y paragraff cyntaf, rwyf fel arfer yn disgwyl gweld rhyw fath o wybodaeth am y cwestiwn. Nawr yn y cynllun marcio, nid oes llawer o farciau am wybodaeth, felly mae'r wybodaeth honno'n hollol iawn.

... gyda [y cynllun marcio] yn fy mhen, rwy'n gwybod nad oes rhaid iddynt gael gormod o wybodaeth.

Wrth farcio, byddai'r cyfranogwyr felly'n rhoi mwy o sylw i agweddau sy'n berthnasol i AA2 ac AA3. Fodd bynnag, mae'n ymddangos bod pwysigrwydd gwybodaeth yn fwy yn ystod BG, ac yn enwedig pan fydd cyfranogwyr yn dod ar draws ymatebion sy'n dangos lefelau tebyg o sgil hanesyddol. Mewn marcio confensiynol, byddai elfen wybodaeth y ddau ymateb wedi cael dau farc (oni bai bod gwallau amlwg). Fodd bynnag, yn ystod BG, mae ymateb yn cael mantais os yw'r beirniaid o'r farn ei fod yn dangos gwybodaeth fwy soffistigedig neu fanylach.

Felly, mae'n ymddangos y gallai BG roi mwy o bwyslais ar wybodaeth gyd-destunol a ddefnyddir i gefnogi'r gwerthusiad. Yn wir, soniodd cyfranogwyr o Grŵp NF am wybodaeth neu ymwybyddiaeth o gyd-destun mewn tua 40% o'u sylwadau, sy'n debyg i gyfran y sylwadau a grybwyllwyd o dan werthusiad. Fodd bynnag, mae'n bosibl wrth farcio fod gwybodaeth hefyd yn cael ei hystyried y tu hwnt i AA1 gan y gellid dadlau na all dysgwyr ddadansoddi a gwerthuso'r ffynonellau heb ddangos gafael ar wybodaeth gyd-destunol.

6.3.3. Cymhlethdod y dadansoddiad

Nododd y cyfranogwyr fod arddangos sgiliau hanesyddol ehangach yn agwedd bwysig arall a ystyriwyd yn ystod BG. Roeddent yn ei ystyried yn enghraifft bellach o ansawdd ymateb yn ogystal ag archwilio dibynadwyedd a defnyddioldeb ffynonellau. Roedd y medr hwn yn cynnwys cymryd y dadleuon am ddibynadwyedd a defnyddioldeb ffynhonnell a'u datblygu ymhellach drwy, er enghraifft, esbonio a ellid cyfyngu'r dadleuon, neu drwy gymharu ffynonellau. Un agwedd arall ar y medr hwn oedd y gallu i integreiddio gwybodaeth mewn gwerthuso a dadansoddi a fyddai hefyd yn arwain at ysgrifennu llai mecanistig. Byddai ymateb datblygedig hefyd yn esbonio canlyniadau'r dadansoddiad, yn tynnu sylw at y ddirnadaeth, ac yn ystyried y cyd-destun ehangach, sydd hefyd yn bwysig ar gyfer dadansoddi dibynadwyedd a defnyddioldeb yn sylfaenol.

Mae soffistigeiddrwydd eithaf braf i hon yn y ffordd y mae'r ymgeisydd yn gwau i mewn. Mae'n dadansoddi'r ffynhonnell ac yna mae'n ei rhoi yn ei chyd-destun gyda'r hyn y mae'n ei wybod ac yna mae'n dod ag ef yn ôl i'r cwestiwn ac yn esbonio pam mae'n ddefnyddiol. Mae ganddi nawws feirniadol braf ynddi hefyd. Felly, mae hi'n dal i feirniadu'r dystiolaeth. Mae'n braf ac yn soffistigedig.

6.3.4. Arddull ysgrifennu fecanistig

Roedd osgoi arddull ysgrifennu fecanistig yn agwedd arall a ystyriwyd gan gyfranogwyr wrth feirniadu. Er ei bod yn ymddangos yn anodd diffinio'n gryno beth oedd arddull ysgrifennu fecanistig, roedd yn ymwneud yn bennaf ag esboniad a oedd yn ceisio ticio'r blychau yn y cynllun marcio. Er enghraifft, byddai gan ymateb

mecanistig bob paragraff yn mynd i'r afael â maen prawf o'r cynllun marcio a/neu byddai'n gorbwysleisio rhai geiriau allweddol (e.e. dibynadwy, defnyddiol). Oherwydd y dull fformiwläig hwn, nid yw'r ymateb yn teimlo fel gwaith integredig ac mae'n ddifflach.

Roedd y cyfranogwyr o'r farn, wrth farcio, fod ymatebion felly'n debygol o gael yr un marciau â'r atebion mwy soffistigedig, oherwydd eu bod yn ticio'r meini prawf marcio. Fodd bynnag, wrth feirniadu, roeddent o dan anfantaes o'u cymharu ag ymatebion a oedd yn hytrach na 'chymryd y camau' yn goresgyn y fframwaith mecanistig hwn ac yn dangos mwy o feddwl.

Mae'r llall yn iawn, ond mae'n amlwg, wyddoch chi, maen nhw'n sôn am ddibynadwyedd ond nid yw'n gwneud unrhyw beth mewn gwirionedd oherwydd mae'n ei wneud oherwydd dywedwyd wrthynt am wneud hynny.

Felly mae'r ddau wedi gwneud Ffynhonnell A3, felly maen nhw wedi sôn am ddefnyddioldeb. Mae'n... arddull ysgrifennu fwy soffistigedig hefyd yn A; mae'n ymddangos yn llawer mwy mecanistig yn B. Wyddoch chi, i raddau, mae'r ffordd y maen nhw'n cyflwyno pethau yn ailadroddus ac yn dilyn strwythur penodol i raddau helaeth, yn hytrach nag yn llo.

Mae [B] yn fwy ymwybodol o bwysigrwydd y geiriau allweddol nag A. Ond mae A yn ... mae'n, [...] gynhesach. Mae'n fwy dynol, wyddoch chi. Byddai'n well gennyf gael sgwrs gyda'r myfyriwr hwn na'r myfyriwr hwn.

6.3.5. Rôl yr elfennau strwythurol

Ar wahân i ofynion cynnwys, roedd o leiaf rai o'r cyfranogwyr yn ystyried elfennau strwythurol ymateb wrth feirniadu. I rai ohonynt, roedd strwythur da sy'n cynnwys 'cyflwyniad, casgliad a phedwar paragraff' yn arwydd y dylai'r ymateb fod yn cydfynd yn fras â'r gofynion ar gyfer traethawd da. Nid yw hyn yn gwrth-ddweud pwynt y cyfranogwyr ar werthfawrogi ysgrifennu nad yw'n fecanistig i'r graddau bod y cysyniad hwn yn cael ei gymhwyso i gynnwys yr ymateb ac nid y strwythur. Mae angen trefnu hyd yn oed ymateb nad yw'n fecanistig er mwyn iddo lifo o un elfen strwythurol i'r llall.

Fy mhrosesau cychwynnol oedd bwrw golwg cyflym a brasddarllen ac edrych ar y strwythur gwirioneddol.

Ystyriwyd bod y cyflwyniad yn ddefnyddiol ar gyfer deall yr hyn y bydd y dysgwr yn ei gynnwys neu'n mynd i'r afael ag ef yn y papur. Wrth farcio, ni fyddai athrawon yn ystyried rhoi pwyntiau ar gyfer y cyflwyniad yn benodol gan nad oes rhaid i ymateb 'gael cyflwyniad' o reidrwydd, ond hefyd oherwydd y byddai cyflwyniad fel arfer yn

cynnwys gwybodaeth generig na fyddai'n cyfrif tuag at farciau'r ymgeisydd. Serch hynny, gallai dysgwyr gollu marciau os yw'r cyflwyniad ar goll neu'n wan, gan y gallai greu argraff negyddol o'r ymateb.

Yn ystod BG, defnyddiodd y cyfranogwyr y cyflwyniad fel maen prawf penderfynu pan oedd dau ymateb o ansawdd tebyg. Soniodd y cyfranogwyr yng Ngrŵp F am y cyflwyniad mewn tua 5% o'r sylwadau. Fodd bynnag, roedd cyfeiriadau at y cyflwyniad yn ddatganiadau cyffredinol, megis 'ffocws clir yn y cyflwyniad'; 'cyflwyniad generig'.

Mae'r casgliad yn elfen strwythurol bwysig, a byddai casgliad gwan neu ddiffyg casgliad yn gostwng y sgoriau wrth farcio. Yn ystod BG, defnyddiwyd y casgliad i greu'r argraff gyffredinol am yr ymateb; i dorri'r ddadl ar gyfer ymatebion o ansawdd tebyg; ac fel agwedd sy'n cael ei hystyried yn fanylach wrth fwrw golwg cyflym neu beidio â darllen ateb yn llawn. Soniodd y cyfranogwyr am y casgliad mewn mwy na chwarter o'u sylwadau.

6.3.6. Pwysigrwydd y geiriau allweddol

Defnyddiodd y cyfranogwyr eiriau allweddol i lunio barn am ansawdd yr ymateb ac i nodi'r testun y byddent yn ei ddarllen yn fwy gofalus wrth fwrw golwg cyflym. Yn gyffredinol, byddai rhai geiriau allweddol yn creu argraff fwy cadarnhaol am yr ymateb.

Gellid categoreiddio'r geiriau allweddol y rhoddodd y cyfranogwyr sylw iddynt yn dri math: geiriau pwnc, geiriau meini prawf asesu, a geiriau trosiannol/dadansoddi.

Mae geiriau pwnc yn benodol i bwnc yr ymateb. Enghreifftiau o'r geiriau hyn yw 'propaganda' a 'chomiwnyddiaeth'. Mae geiriau meini prawf asesu, megis 'dibynadwy', 'rhagfarnllyd' a 'defnyddiol' yn berthnasol i amcanion asesu. Defnyddiwyd y ddau grŵp hyn o eiriau i nodi'r darnau yr oedd angen rhoi mwy o sylw iddynt wrth fwrw golwg cyflym.

Mae'r trydydd math o eiriau allweddol, megis 'fodd bynnag', 'mae hyn yn awgrymu' yn tynnu sylw at elfennau dadansoddol yr ymateb. Er enghraifft, pan fydd dysgwyr yn defnyddio math o frawddeg 'mae hyn yn awgrymu', byddai'n dangos i'r beirniaid fod y dysgwyr hynny'n 'darllen i'r ffynhonnell' ac yn meddwl am yr hyn y mae'r ffynhonnell yn ei olygu. Mae'r geiriau hyn hefyd yn bwysig wrth farcio, er nad ydynt o reidrwydd yn cael eu hadlewyrchu yn y cynllun marcio.

Mae'n sôn am ddefnyddiol, defnyddiol, defnyddiol. Mae'n ddefnyddiol o ran gwerthuso, yn ddefnyddiol iawn, hefyd yn ddefnyddiol, dyna ni. ... Ydy, iawn, mae'n amleiriog ond mae'n gwneud yr hyn mae'n ei ddweud ar y tun.

[...] mae wedi defnyddio'r gair [dibynadwy]; ond a yw wedi'i asesu mewn gwirionedd? Mae'n ddigon posibl ei fod yn rhagfarnllyd ... Iawn, felly mae'n amwys iawn, iawn, [...] heb ei archwilio'n iawn mewn gwirionedd.

Rwy'n chwilio am 'fodd bynnag', sef, wyddoch chi, pan fydd myfyriwr yn edrych ar y cydbwysedd, ac nid oes 'fodd bynnag' yn hyn o gwbl.

Nid wyf wedi dod o hyd i lawer o [eiriau allweddol] yn A, felly mae'n gwneud i mi feddwl mai ymgeisydd B yw, wn i ddim, y gorau, gallech ei ddweud.

Cyn gynted ag y gwelaf y geiriau hynny, rwy'n canolbwyntio ar y brawddegau hynny'n amlwg i wirio'r sylwadau dibynadwyedd a'r sylwadau defnyddioldeb a sut y cânt eu cymhwyso i'r atebion.

6.4. Agweddau heriol ar feirniadu

Y rhan fwyaf o'r amser ni chafodd y cyfranogwyr unrhyw anawsterau wrth benderfynu p'un o'r ddau ymateb oedd yn well. Fodd bynnag, mewn cyfran fach o farnau, canfu'r cyfranogwyr fod angen iddynt ystyried mwy o agweddau ar yr ymatebion cyn penderfynu. Digwyddodd hyn yn rhannol oherwydd blinder naturiol, ond hefyd am fod yr ymatebion yn cyflwyno rhai anawsterau a drafodir yn yr isadrannau canlynol.

6.4.1. Beirniadu ymatebion o werth tebyg

Teimlai'r cyfranogwyr fod yr ymatebion pâr, yn ail hanner y sesiwn feirniadu, o werth mwy tebyg, sy'n cyd-fynd â sut y byddai algorithm BG yn paru'r ymatebion. Roedd angen ystyried yr ymatebion hyn yn fwy gofalus ac roeddent yn anos eu penderfynu o ran ansawdd cymharol. O'r ymatebion hyn, roedd y cyfranogwyr o'r farn ei bod yn arbennig o anodd gwahaniaethu rhwng dau ymateb gwannach oherwydd eu bod 'i gyd yn dweud yr un peth' neu'n 'swnio'r un fath'. Ar yr un pryd, byddai gan ymatebion cryf bob amser rywbeth a fyddai'n helpu sgorwyr i wahaniaethu'n haws rhyngddynt.

Teimlwyd bod gorfod cymharu a phenderfynu ar enillydd rhwng ymatebion o werth tebyg 'ychydig yn artiffisial' oherwydd wrth farcio, byddai'r ddau ymateb yn cael yr un marciau. Yn ystod BG, mabwysiadodd y cyfranogwyr strategaethau amrywiol i wneud y penderfyniadau hyn.

Byddai rhai cyfranogwyr yn ailedrych ar rai elfennau o ymateb, megis y testun o amgylch y geiriau allweddol, casgliad neu 'sbardunau' eraill yr amcanion asesu.

Pan fydd dau yn eithaf tebyg i'w gilydd, yna byddwn yn meddwl, "Beth ydw i'n chwilio amdano mewn gwirionedd?" A dw i'n meddwl ... ac yn yr achosion hynny, roeddwn yn edrych i weld a oeddent wedi deall rhagfarn y ffynhonnell.

Roeddwn i'n canolbwyntio i weld a ydy un ychydig yn fwy soffistigedig na'r llall, neu a ydy'r naill wedi cynnwys y gynulleidfa a'r llall heb?

Nid oedd y cyfranogwyr eraill yn canolbwyntio ar un agwedd ond yn ceisio dod o hyd i unrhyw wahaniaethau cynnil rhwng y ddau ymateb.

Felly, efallai bod yn rhywbeth a oedd efallai mewn un ... gallai fod wedi bod yn rhywbeth mor syml â sillafu, atalnodi a gramadeg. Ac rydym yn marcio hyn hefyd oherwydd AA1, mae'n ymwneud â chydlynid, ansawdd cyfathrebu ysgrifenedig a dealltwriaeth o wybodaeth.

Lawer gwaith, byddai cyfranogwyr yn dibynnu ar eu greddf i benderfynu rhwng dau ymateb tebyg. Roedd y rhan fwyaf o'r cyfranogwyr yn agored yn ei gylch, tra bod rhai'n fwy tawedog yn ei gydnabod.

Rhodddwyd cyfarwyddyd i'r cyfranogwyr y gallent ddewis ymateb ar hap pan nad ydynt yn siŵr pa ymateb sy'n well (e.e. drwy fflipio darn arian), ond nid ydym yn ymwybodol a wnaeth unrhyw un o'r cyfranogwyr hynny. Soniodd y cyfranogwyr hynny y gofynnwyd iddynt am y dull hwn yn ystod cyfweiliadau ei bod yn well ganddynt benderfynu drostynt eu hunain na'i adael i siawns. Yn ddamcaniaethol, gallai hyn fod yn broblem gan y byddai'r model yn tybio penderfyniad ar hap pan fydd gan y sgrïptiau yr un gwerth ar y priodoledd cudd. Gallai peidio â phenderfynu ar hap greu rhagfarn gan ddibynnu ar yr agwedd a ddefnyddir i benderfynu ar yr enillydd. Yn ein hastudiaeth, ychydig iawn o achosion o'r fath a grybwyllwyd.

6.4.2. Beirniadu ymatebion ar wahanol bynciau

Roedd rhai cyfranogwyr yn ei chael hi'n anodd beirniadu ymatebion ar wahanol bynciau, neu'n meddwl y gallai rhai athrawon ei chael hi'n anodd beirniadu ymatebion o'r fath.

Felly, roeddwn yn teimlo'n gyson nad oeddwn yn mynd i fod 100% yn gywir [yn beirniadu] y rheini. Er eu bod i gyd yn amodau cymdeithasol o'r 19eg ganrif, neu rwy'n credu mai hwinnw a wnaethom ar gyfer y llall, rwy'n teimlo y byddwn wedi bod yn llawer mwy cywir pe byddwn yn gwneud yr un peth yn gyson.

Teimlai'r cyfranogwyr y gall gwybod mwy am un pwnc na'r llall effeithio ar gywirdeb y farn, gan nad yw beirniaid efallai'n sicr o ansawdd yr ymateb ar y pwnc llai cyfarwydd. Gallai hyn wneud iddynt asesu'r ymateb ar y pynciau anghyfarwydd naill ai'n fwy llym neu'n fwy trugarog (gan ganiatáu ar gyfer mantais yr amheuaeth).

Felly, gallai fod yn broblem os nad yw'r person sy'n marcio yn addysgu'r maes hwnnw mewn gwirionedd, heb fod yn gwybod llawer amdano. Felly sut ydych chi'n gwybod beth maen nhw'n ei ddweud am amodau Dwyrain Llundain? Sut ydych chi'n gwybod bod hynny'n ffeithiol gywir?

[Cyn y gweithdy] Edrychais ar y ffynhonnell ac ar y cefndir, ond efallai fod hynny braidd yn heriol. [...] Credaf y byddai'n ddefnyddiol iawn pe bai'r person yn addysgu'r pwnc penodol hwnnw.

Efallai y bydd athrawon hefyd yn mwynhau un pwnc yn fwy na'r llall, a gall hyn greu tuedd anymwybodol tuag at ffafrio ymatebion ar y pwnc a ffafirir. Fel yr eglurodd un cyfranogwr, mewn achosion o'r fath, gall beirniaid deimlo'n dda am ymateb, ond gallai hyn fod am y rhesymau anghywir.

Er gwaethaf yr anawsterau hyn, roedd cyfranogwyr yn dal i ystyried y gallai BG weithio'n dda hyd yn oed wrth feirniadu ymatebion ar wahanol bynciau. Teimlwyd y gallai cael ymatebion ar wahanol bynciau hyd yn oed wneud i feirniaid ganolbwyntio mwy ar fedr y dysgwyr yn hytrach na chael eu llethu gan y cynnwys.

Fodd bynnag, byddai'r barnau hyn yn fwy heriol yn wybyddol, yn gofyn am fwy o amser ac mae'n debyg eu bod yn llai dibynadwy. Er mwyn gwneud beirniadu ymatebion ar wahanol bynciau yn fwy cywir, awgrymwyd bod athrawon yn cael rhywfaint o wybodaeth gefndir am y pwnc anghyfarwydd cyn beirniadu.

Er mwyn gwneud BG yn ddichonadwy, byddai'n ofynnol wedyn bod pynciau'r NEA wedi'u cyfyngu i nifer rhesymol⁶. Yn gyffredinol, nid oedd y cyfranogwyr yn erbyn y syniad hwn. Dywedodd rhai na fyddai ots ganddynt pe bai ysgolion yn astudio'r un pwnc, gan eu bod yn teimlo y byddai hyn yn sicrhau maes cystadlu cyfartal i'r asesiad ac yn gwneud canlyniadau'r asesiad yn decach. Fodd bynnag, roedd cyfranogwyr eraill yn anghytuno â hyn oherwydd byddai o fantais i'r ysgolion sydd â mwy o brofiad gyda'r pwnc. Roedd y cyfranogwyr hyn yn cefnogi'r syniad o nifer cyfyngedig

⁶ Un dull arall fyddai gofyn i athrawon gymharu gwaith eu dysgwyr wedi'i gymysgu â chyfran lai (e.e. 20%) o barau gan ddefnyddio sgrïptiau o ysgolion eraill. Mae'r sgrïptiau o ysgolion eraill megis angorau a defnyddir eu gwerthoedd amcangyfrifedig i raddnodi gwerthoedd yr holl sgrïptiau dan sylw mewn BG (gweler Wheadon, C., et al., 2020a). Anfantais y dull hwn o'i gymharu â'r dull a ddefnyddir yn yr astudiaeth hon yw nad yw'n dileu rhywfaint o ragfarn yr athro ac nad yw'n cael effaith debyg o ran datblygiad proffesiynol, mae'n debyg.

o bynciau yn unig, gan y byddai hyn yn caniatáu i ysgolion ddewis pwnc yn nes at eu harbenigedd ac anghenion y dysgwyr.

Credaf pe bai pawb yn gwneud rhywbeth tebyg, gallech fonitro o ran y cymorth a'r mynediad at y ffynonellau y dylent fod yn eu canfod yn annibynnol yn hawdd.

Byddwn yn pryderu am y disgyblion hynny sydd wedi dilyn Safon Uwch oherwydd bod ganddynt frwdfrydedd naturiol dros rywbeth ac efallai eu bod am wneud pwnc nad yw o reidrwydd o fewn y maes llafur. Gallent gael eu cosbi wedyn am hynny.

Yn gyffredinol, cytunodd y cyfranogwyr y byddai'n anos ac yn llai effeithlon cymharu ymatebion ar wahanol bynciau. Roeddent yn credu y byddai hyd yn oed yn anos cymharu ymatebion rhan (b) gan fod y rheini'n fwy cymhleth, a bod mwy o angen byth i ystyried yr hanesyddiaeth. Fodd bynnag, byddai angen mwy o ymchwil i werthuso'r safbwyntiau hyn yn feintiol.

Fel gyda'r ymatebion gwerth tebyg, teimlwyd bod ymatebion gwannach ar wahanol bynciau yn anoddach eu beirniadu am fod llai o dystiolaeth o sgiliau uwch, a bod gwybodaeth hanesyddol yn dod yn bwysicach i lunio'r farn. Fodd bynnag, ni fyddai'r beirniaid yn gallu ei gwerthfawrogi gan nad oedd y pwnc yn gyfarwydd iddynt.

6.4.3. Beirniadu ymatebion ysgrifenedig

Nid yw'n syndod bod cyfranogwyr wedi sôn ei bod yn llawer haws darllen ymatebion wedi'u teipio na'r rhai llawysgrifen, yn enwedig y rhai â llawysgrifen lai darllenadwy. Roedd gorfod darllen ymatebion llawysgrifen wedi'u sganio ar sgrin yn gwneud y darllen yn anoddach byth. Ond nid oedd hyn fel petai'n broblem a gafodd effaith fawr ar farnu, gan nad oedd mwyafrif helaeth o'r cyfranogwyr yn teimlo unrhyw anghyfleustra yn y broses o lunio barn.

Nid oedd gorfod darllen ymatebion llawysgrifen fel petai'n effeithio ar y farn am yr ymatebion hynny. Fel y soniwyd yn gynharach, gofynnwyd i'r cyfranogwyr raddio darllenadwyaeth yr ymatebion ysgrifenedig. Nid oedd graddio darllenadwyaeth ymateb yn cydberthyn â threfn restrol BG yn yr un o'r ddau grŵp.

Ond nid yw hyn yn golygu nad yw athrawon yn talu sylw i ansawdd llawysgrifen o gwbl. Yn ystod y cyfweiliadau, er enghraifft, gwnaeth athrawon sylwadau ar lawysgrifen a'i nodweddu fel 'taclus' neu 'anaeddfed'. Weithiau cyfeiriodd yr athrawon at y dysgwr gan ddefnyddio'r rhagenw personol 'hi', gan gasglu'r rhyw o daclusrwydd y llawysgrifen.

Gadewch i ni ddechrau gyda hwn. A dw i'n cofio hwn [o'r gweithdy]. 'Achos dw i'n cofio dweud, "Rhaid mai merch ydyw, rydych chi'n gallu dweud." Byddai'n ddiddorol pe gallech chi wneud llawysgrifen merch neu fachgen ac rwy'n mentro dweud y byddem yn ei gael yn iawn.

Yn gyffredinol, soniodd y cyfranogwyr y byddai'n well beirniadu ymatebion wedi'u teipio. Fodd bynnag, roeddent yn amau y gallai pob ysgol sicrhau mynediad i gyfrifiaduron. Roedd rhai hefyd yn poeni a allai dysgwyr deipio eu hymatebion yn ddigon cyflym yng nghyd-destun asesiad amser cyfyngedig.

Mae gennym broblem enfawr gyda mynediad i gyfrifiaduron i fyfyrwyr nad ydynt yn gwneud TG, Dylunio a Thechnoleg a phynciau y mae angen cyfrifiaduron arnynt.

6.4.4. Gwneud sylwadau

Fel y soniwyd yn gynharach, gofynnwyd i hanner y beirniaid roi adborth neu roi sylwadau ar yr ymatebion.

Gwnaeth y cyfranogwyr 204 o sylwadau i gyd, gan wneud 9 sylw ar gyfartaledd fesul ymateb a 18.5 sylw fesul beirniad, sy'n golygu bod beirniaid wedi gadael tuag un sylw fesul barn a wnaethant. O gyfanswm y sylwadau, gwnaed 20 yn ogystal â'r rhai cychwynnol. Roedd y sylwadau'n amrywio o ddau air i 64 gair. Roedd sylwadau byrrach yn fwy cyffredinol ac mae'n debyg eu bod yn llai defnyddiol yn wahanol i sylwadau manylach, fel y gwelir o'r sylwadau isod a wnaed ar dri ymateb gwahanol:

Cyfranogwr 1: Cyflwyniad gwan

Cyfranogwr 2: Yn gwerthuso ffynonellau

Cyfranogwr 3: Cyflwyniad mwy cyffredinol, er bod iddo ffocws. Yn dadansoddi dyfyniadau ac yn darparu manylion hanesyddol sy'n berthnasol. Mae'r dadansoddiad o ddefnyddioldeb ychydig yn fformiwlâig e.e. cynnwys yna tarddiad. Rhai pwyntiau gor-symbl e.e. gallai'r Daily Mail fod yn rhagfarnllyd gan nad oedd yn Americanaidd. Mae'r casgliad ar y cwestiwn.

Dywedodd y cyfranogwyr fod y sylwadau a wnaethant i fod i'w helpu i feirniadu ac nad oeddent wedi'u bwriadu fel adborth i'r dysgwyr, ac eithrio un cyfranogwr a aeth i'r afael â'i sylwadau ar ffurf adborth i'r dysgwr. Er enghraifft:

Wrth ddadansoddi dibynadwyedd neu ddefnyddioldeb ffynonellau, esboniwyd beth mae hyn yn ei olygu i'ch barn. Ystyriwyd ddiben y ffynonellau rydych chi'n ymchwilio iddynt

Teimlai'r cyfranogwyr fod gwneud sylwadau yn haws nag anodi yn ystod marcio confensiynol. Y rheswm am hyn yw bod anodi confensiynol yn fanylach, ac ar gyfer BG, gadawodd y cyfranogwyr sylwadau mwy cyfannol a lefel uchel.

Er bod rhai cyfranogwyr o'r farn bod anodi wrth farcio yn ddefnyddiol, nid oedd eraill yn teimlo ei fod yn ddefnyddiol iddynt ac roeddent yn teimlo ei fod yn gwneud iddynt dreulio mwy o amser gyda marcio. I'r gwrthwyneb, roedd yn llawer gwell gan gyfranogwyr o'r ddau grŵp wneud sylwadau oherwydd eu bod yn credu y byddai gwneud sylwadau yn helpu i benderfynu wrth gymharu'r ymatebion, yn enwedig wrth ddod ar draws yr ymateb y tro nesaf⁷.

Yn wir, roedd y cyfranogwyr a wnaeth sylwadau yn aml yn defnyddio eu sylwadau i dorri'r ddadl wrth edrych ar yr ymateb yr eildro a phan oedd y sgript arall o ansawdd tebyg. Weithiau, roeddent yn seilio eu penderfyniad ar y sylwadau yn unig yn bennaf, gan frasddarllen y sgript yn gyflym i'w gadarnhau (neu i'r gwrthwyneb). Teimlai rhai cyfranogwyr, weithiau, y byddai dim ond darllen y sylwadau yn ddigon i lunio'r farn.

Pe bawn i wedi ysgrifennu arno 'papur rhagorol', [ac], os oedd y papur arall yna'n warthus, yn wael iawn yna fyddwn i ddim yn ei ddarllen eto.

Ar ôl i chi weld [ymateb], gallech edrych ar y sylwadau [a wnaed am yr ymateb] i'ch atgoffa amdano. Yna nid oedd yn rhaid i chi ei ddarllen eto oni bai bod enghraifft B yn eithaf tebyg, ac yna bu'n rhaid i chi edrych arnynt eto ac yna penderfynu.

Felly, roeddwn yn edrych ar y sylwadau i wirio ddwywaith, ac weithiau roedd gen i sylwadau ar y ddau ohonynt, a byddwn yn mynd yn ôl i frasddarllen yn gyflym dim ond i sicrhau bod fy sylwadau'n gywir. Ac felly, o'r gronfa sylwadau wedyn, roeddwn, wel, yn gallu penderfynu pa un oedd y gorau.

Roedd cyfranogwyr yn credu'n gryf bod gwneud sylwadau yn cyflymu'r beirniadu. Roedd hyd yn oed athrawon na wnaeth sylwadau yn rhannu'r farn hon, gan y gallent sylwi nad oedd y cydweithwyr o'r grŵp arall y tu ôl iddynt gyda nifer y barnau a gyflawnwyd. Cadarnhawyd y canfyddiad goddrychol hwn gan yr amser beirniadu cyfartalog ym mhob grŵp. Credid bod gwneud sylwadau yn arbennig o ddefnyddiol ar gyfer ymatebion hir, er enghraifft, ymatebion rhan (b) a all fod mor hir â 18 tudalen.

⁷ Fodd bynnag, dylem nodi na fyddai hyn yn wir pe bai BG yn cael ei gweithredu ar raddfa fwy, oherwydd gallai fod yn llai posibl dod ar draws yr un ymateb.

Roedd rhai cyfranogwyr (yn enwedig yng Ngrŵp NF) yn poeni y byddai'r sylwadau presennol yn atal beirniaid rhag darllen sgrïpt yn iawn o ddod ar ei thraws y tro nesaf, gan effeithio ar eu cywirdeb. I'r gwrthwyneb, teimlai rhai cyfranogwyr o Grŵp F fod sylwadau darllen yn gwneud eu penderfyniadau'n fwy cywir. Fodd bynnag, roedd pryder y gallai beirniaid feddwl bod sylw a wnaed yn flaenorol yn cyfeirio at yr ymateb yr oeddent yn ei adolygu ar y pryd, ond mewn gwirionedd roedd yn sylw yn cyfeirio at ymateb gwahanol.

Efallai y byddwch yn gweld y sylwadau ac yn dweud, "lawn, dyna'r un," ond yna efallai mai'r llall ydyw.

Mae'r arsylwadau hyn yn awgrymu y gall gwneud sylwadau ychwanegu at amrywioldeb y barnau mewn o leiaf ddwy ffordd. Yn gyntaf, gall gwneud sylwadau effeithio ar y broses farnu ac arwain at ganlyniadau gwahanol o gymharu â phan na wneir sylwadau. Mae'n bosibl, er enghraifft, fod agweddau sydd wedi'u cynnwys yn y sylwadau yn sbarduno rhagfarn argaeledd pan wneir y penderfyniad terfynol, gan ystumio barn gyfannol yr ymarfer. Yn ail, gallai dibynnu ar sylwadau blaenorol effeithio ar ragdybiaeth annibyniaeth y model ystadegol a ddefnyddir i ddadansoddi'r data. O ystyried bod beirniaid yng Ngrŵp F yn teimlo bod gwneud sylwadau'n ddefnyddiol, ac nad oedd yn arafu'r beirniadu, mae'n ymddangos y gallai gwneud sylwadau helpu i feirniadu ymatebion ysgrifenedig hirach yn gyflymach. Fodd bynnag, mae'n debyg nad yw'r fantais hon yn bodoli pan gymhwysir BG ar raddfa fawr, ac os felly mae'r tebygolrwydd o ddod ar draws sgrïptiau blaenorol yn isel iawn. Serch hynny, dylid ymchwilio ymhellach i effeithiau gwneud sylwadau pan fydd angen beirniadu cronfa fach o ymatebion.

Ar wahân i fod yn ddefnyddiol yn ystod BG, gallai sylwadau fod â gwerth ffurfiannol os ydynt yn cael eu trosglwyddo i'r athro,(e.e. i wella gwaith cwrs yn y dyfodol). Fodd bynnag, er mwyn i hyn ddigwydd, byddai angen i sylwadau fod yn fwy penodol a manwl, yn hytrach na chynnwys ychydig o eiriau disgrifiadol yn gyffredinol. Gellid defnyddio sylwadau hefyd i archwilio'r sgrïptiau *misfitting* neu wrth archwilio apeliadau yn erbyn canlyniadau BG.

Awgrymodd y cyfranogwyr y gellid cyflymu'r broses o wneud sylwadau ymhellach drwy gynnwys sylwadau parod i'w defnyddio yn y feddalwedd (e.e. 'gwybodaeth fanwl'). Roedd rhai o'r farn bod angen anodi ymateb ac awgrymasant y gallai'r feddalwedd sicrhau bod codau byr ar gael a fyddai'n debyg i anodi, e.e. A ar gyfer ailadrodd. Teimlai'r cyfranogwyr hefyd y byddai'n ddefnyddiol gallu amlygu testun (e.e. defnyddio lliwiau penodol ar gyfer gwahanol amcanion asesu) ac, os yw'n bosibl, integreiddio sylwadau yng nghorff y papur.

6.5. Cwestiwn BG

Gall y cwestiwn a ofynnir gan feirniaid gyflwyno rhai heriau os nad yw'n glir. Roedd y mwyafrif o'r farn bod y cwestiwn yn gofyn - *Pa sgript sy'n cynrychioli disgybl sy'n well am Hanes?* - a bod hwnnw'n unol â'r hyn y byddent yn edrych arno wrth farcio.

Yr un a fyddai'n well am hanes yn fy marn i fyddai'r un a fyddai'n cael y marciau mwyaf am ddibynadwyedd, defnyddioldeb. Felly, mae'n ... roedd yn gweithio.

Fodd bynnag, credai rhai cyfranogwyr y gellid ei ddehongli'n wahanol sef cyfeirio at gynnwys agweddau eraill heblaw'r rhai a ystyriwyd wrth farcio. Awgrymwyd ganddynt y dylid geirio'r cwestiwn fel ei fod yn nes at y cwestiwn y mae'r dysgwyr yn ei ateb yn eu traethodau, e.e. 'Pa ddysgwr sy'n well am edrych ar ddefnyddioldeb a dibynadwyedd ffynonellau?'.

7. Manteision canfyddedig BG

7.1. Effeithiau ar yr asesiad

7.1.1. Effeithiau ar ddilysrwydd a chreadigrwydd yr ymatebion

Mae astudiaethau ymchwil (gweler Wheadon et al., 2020a) wedi canfod y gallai defnyddio cynlluniau marcio ar gyfer ysgrifennu, er y bwriedir iddynt gynyddu dibynadwyedd asesu, hefyd effeithio'n negyddol ar ddilysrwydd a chreadigrwydd dysgwyr, wrth iddynt ddechrau targedu'r cynllun marcio yn eu hysgrifennu.

Credai'r cyfranogwyr y byddai defnyddio BG wrth farcio yn y pen draw yn llywio dysgwyr i fod yn fwy creadigol ac i ganolbwyntio ar sgiliau a dealltwriaeth yn hytrach nag ar y 'gallu i neidio drwy gylchoedd'. Y rheswm am hyn yw bod BG yn gofyn am ddull cyfannol o asesu, a bod gwerth ymateb yn cael ei ddiffinio gan ansawdd cymharol ymatebion eraill. Roeddent o'r farn, er y byddai'r llyw hwn yn cymhell y dysgwyr pen uchel i ysgrifennu atebion gwell, na fyddai'r dysgwyr llai galluog o dan anfantais, oherwydd ni ddylai ysgrifennu ymatebion fod yn anos nag y mae ar hyn o bryd. Awgrymwyd y byddai'r newid hwn yn cynnig gwell paratoad ar gyfer dilyniant pellach o ran hyrwyddo creadigrwydd wrth agosáu at eu hasesiadau.

Efallai y byddai myfyrwyr yn cael eu gwobrwyo'n well am eu hunigolrwydd a'u dawn nag y maent efallai os ydynt wedi'u marcio yn erbyn y cynllun marcio, a'r cynllun marcio yn unig. Oherwydd os ydynt yn marcio yn erbyn y cynllun marcio yn unig, yna nid ydynt wedi gwneud rhywbeth y mae'r cynllun marcio'n dweud bod angen iddynt ei wneud, gall fod mor wych ag yr hoffech ond ni

fyddant yn cael y marciau amdano o hyd, ond mae hyn yn caniatáu ichi wobrwyo'r ddawn a'r unigolrwydd hwnnw'n fwy efallai.

Teimlai'r cyfranogwyr y byddai BG hefyd yn annog arddull addysgu sy'n canolbwyntio mwy ar y sgiliau nag ar yr elfennau y mae angen eu cynnwys yn eu hymatebion. Nid yw pwyslais gwybodaeth gywir am ddigwyddiadau hanesyddol gan rai cyfranogwyr wrth feirniadu o reidrwydd yn gwrth-ddweud hyn, gan y gellid ystyried ei fod yn rhan annatod o fedr. Roedd rhai cyfranogwyr o'r farn, pe bai BG yn cael ei defnyddio wrth farcio, y byddai'n well ganddynt ganolbwyntio ar atebion enghreifftiol (gan gynnwys papurau Safon Uwch ac israddedig) ac addysgu sgiliau hanesyddol ohonynt. Awgrymodd rhai y gallent hyd yn oed ddefnyddio BG fel offeryn ffurfiannol i'r dysgwyr drwy ofyn iddynt gymharu gwahanol ymatebion.

7.1.2. Tegwch

Roedd cyfranogwyr o'r farn bod athrawon a oedd yn marcio eu dysgwyr eu hunain yn dueddol o fod rhywfaint yn oddrychol. Mae hyn oherwydd bod syniadau rhagdybiedig athrawon am alluoedd eu dysgwyr yn gwneud marcio gwrthrychol yn anodd. Nodwyd hefyd ei bod yn debygol bod athrawon yn 'buddsoddi'n emosiynol' yn eu dysgwyr eu hunain ac felly eu bod yn fwy trugarog wrth eu marcio.

Byddai [BG] yn gwaredu rhywfaint o oddrychedd oherwydd, rwy'n credu ... p'un a ydym yn ei hoffi ai peidio, mae tuedd bob amser tuag at ein dysgwyr. Ac rydych chi bob amser yn ceisio bod, byddwn i'n onest, mor oddefgar â phosibl.

Roedd rhai cyfranogwyr hefyd yn credu efallai na fyddai dulliau marcio athrawon yn gyson ar draws ysgolion. Er enghraifft, roedd cyfranogwyr yn teimlo bod amharodrwydd ymhlith rhai athrawon i roi marc sero neu hyd yn oed un marc ar gyfer amcan asesu ar wahân oherwydd gellid ystyried bod hyn yn 'gas' neu heb fod yn cefnogi ymdrech y dysgwr. I'r gwrthwyneb, byddai rhai athrawon yn fwy amharod i roi marciau llawn na'r lleill. Gallai'r rhagduddiadau hyn wahaniaethu'n annheg yn erbyn dysgwyr mewn rhai ysgolion, yn enwedig pan fo ystod gyfyngedig o farciau ar gael, fel sydd ar gyfer ymateb rhan (a).

Credai'r cyfranogwyr y byddai asesu ymatebion NEA gan ddefnyddio BG yn gwneud asesiad a gynhaliwyd mewn canolfannau yn fwy gwrthrychol na defnyddio marcio confensiynol, gan y byddai'n dileu rhai rhagfarnau o farcio confensiynol. Er enghraifft, roeddent o'r farn bod cymharu dau ymateb yn dileu'r duedd drugarog neu lem y gallai rhai marcwyr ei harddangos wrth farcio. Byddai cywirdeb yr asesiad hefyd yn cynyddu oherwydd byddai llawer mwy o athrawon yn ystyried yr un ymateb.

Rwyf yn ei farcio unwaith, rwyf yn mynd ag ef i'w gymedroli, efallai y caiff ei gymedroli, efallai na fydd. Felly, mae'r sgrïpt honno'n mynd i mewn. Wedi gwneud. Heb ei gweld byth gan gydweithiwr arall.

O ran rhagfarn, mae'n werth sôn bod posibilrwydd, os mai dim ond ychydig o'r holl feirniaid sy'n llunio barn yn y cylchoedd cyntaf, y gallent ogwyddo'r broses ymaddasolrwydd pe baent yn rhagfarnu mewn unrhyw ffordd (e. e. roeddent i gyd yn rhoi'r un pwysiad anghymesur ar agwedd benodol wrth feirniadu ymatebion, e.e. taclusrwydd llawysgrifen). Yn ein hastudiaeth, dechreuodd pob beirniad feirniadu ar yr un pryd. Fodd bynnag, o'r 46 barn yn y pedwar cylch cyntaf yng Ngrŵp NF, gwnaed tua 57% ohonynt gan dri beirniad a 67% gan bedwar beirniad (allan o 12 beirniad i gyd). Yng Ngrŵp F, o'r un nifer o farnau a beirniaid, gwnaed 50% o farnau gan dri beirniad a 61% ohonynt gan bedwar beirniad. Ni allwn ddweud a oedd hyn yn gogwyddo'r canlyniadau yn ein hastudiaeth. Fodd bynnag, mae'r nifer anghymesur o farnau a wnaed gan bob beirniad yn y cylchoedd cychwynnol hefyd yn awgrymu y byddai'n anodd iawn sicrhau y byddai pob beirniad yn cynnal niferoedd tebyg o farnau cychwynnol wrth gymhwyso BG ar raddfa fawr os nad oes cyfyngiadau ar waith. Un ateb fyddai cynllunio'r algorithm didoli fel nad yw'n caniatáu i'r beirniaid mwy gweithredol fynd y tu hwnt i'r rowndiau cyntaf nes bod pob beirniad wedi cymryd rhan.

Yn olaf, soniodd sawl cyfranogwr y gallai BG leihau posibilrwydd twyllo. Soniwyd ei bod yn bosibl y gallai athrawon roi mwy o farciau i bob ymateb ac yna gofyn i ddysgwyr y dewiswyd eu gwaith i'w safoni'n allanol gywiro neu ychwanegu at eu hymatebion. Ni ddywedodd y cyfranogwyr eu bod yn gwybod bod hyn wedi digwydd, ond roedd yn ymddangos bod y posibilrwydd pur fod hyn yn bosibl yn peri pryder iddynt. Er na fyddai BG yn atal hyn yn llwyr, credid y gallai ei wneud yn llai tebygol, oherwydd ni fyddai athrawon yn teimlo bod angen iddynt amddiffyn marc mewn craffu allanol. Yn ogystal, mae uwchlwytho'r ymatebion yn fuan ar ôl eu cwblhau yn creu llai o bosibiliadau ail-weithio amhriodol.

7.1.3. Rhwyddineb gwneud penderfyniadau

Roedd cyfranogwyr o'r farn bod cymharu ymatebion yn haws na neilltuo marc penodol i bob ymateb.

Pan fyddwch chi'n dechrau, yn marcio asesiad, mae'n anodd iawn a dydw i ddim yn meddwl eich bod byth mewn hyfforddiant athrawon yn cael syniad o sut i farcio. Ond rwy'n credu y gall pawb feirniadu. Felly, rwy'n meddwl ar gyfer athrawon mwy newydd ac i'r rhan fwyaf o bobl, gallwch farnu'r gwahaniaeth rhwng rhywbeth sy'n well neu'n waeth.

Pe bai gen i ddechreuwr newydd, byddai'n well gen i petaent yn cymharu'r rhain yn hytrach na cheisio rhoi marc penodol ar rywbeth. Credaf fod hynny'n llawer anoddach.

Soniodd y cyfranogwyr fod rhwyddineb cymharu yn arbennig o ddefnyddiol ar gyfer pynciau fel hanes, yr oeddent yn eu hystyried yn oddrychol iawn ac felly'n anodd eu marcio'n gywir. Mae'r canfyddiad hwn o oddrychedd yn adleisio'r canfyddiadau mai cwestiynau sydd wedi'u hanelu at archwilio ffynonellau yw'r rhai anoddaf i'w marcio mewn arholiadau Hanes oherwydd posibilïadau ymatebion a dehongliadau mor amrywiol o'r ffynonellau (Holmes et al., 2018, t.12).

Gallech gael tri chydweithiwr mewn ysgol sydd i gyd yn marcio set ac yna'n dod at ei gilydd. Maen nhw wedi ei haddysgu yn yr un ffordd; maen nhw wedi'i marcio yn yr un ffordd. Ac maen nhw'n eistedd i gymedroli, a dydyn nhw ddim yn cytuno.

Ond yna mae'n rhaid i chi geisio dod i gytundeb yn erbyn y cynllun marcio o ran pa farc rydych chi'n mynd i'w roi iddo. Ac rydych chi'n pwysu a mesur, "O, ai 11 yw hwnnw neu 12? Wn i ddim." Ond gyda hyn, rwy'n credu ei bod yn haws i mi ddweud, "Na, mae hwnnw'n well na hwnnw."

Mae anhawster sicrhau marcio cywir yn rhoi baich seicolegol ychwanegol ar athrawon oherwydd eu bod yn teimlo, o ystyried yr ystod gul o farciau sydd ar gael, ei bod yn fwy tebygol y gallai gwall un marc wneud gwahaniaeth i ddysgwyr. Er mwyn lleihau'r pwysau seicolegol hwn, dywedodd y cyfranogwyr fod yn rhaid iddynt weithiau ystyried ymateb yn helaeth neu hyd yn oed, fel y dywedodd un cyfranogwr, 'ymboeni' ynghylch y marciau cywir.

Oherwydd dyna lle rwy'n cael marcio gwaith cwrs yn straen, pan fyddwch yn y cyfyng-gyngor hwnnw, "Ydw i'n rhoi hwn iddo? Ydw i'n rhoi hwnnw iddo? Pa farc ydw i'n mynd i'w roi iddo?"

Teimlai'r cyfranogwyr y byddai BG yn dileu'r pwysau seicolegol hwn, nid yn unig am fod cymharu'n haws ond hefyd oherwydd y byddai llawer mwy o athrawon yn asesu'r un ymateb, sy'n gwneud camgymeriad yn llai tebygol. Gobeithio y dylai hyn olygu bod pob sgript yn cael 'ei lle cywir a'i marc cywir'.

Soniodd y cyfranogwyr, ar wahân i bwysu a mesur faint o farciau yw gwerth ymateb, y byddent hefyd yn cymharu gwahanol ymatebion. Mae hyn yn eu helpu i raddnodi eu marcio ac yn sicrhau bod marcio'n deg yn gyffredinol. Weithiau mae hyn yn golygu bod yn rhaid iddynt fynd yn ôl at bapurau sydd eisoes wedi'u marcio i sicrhau bod y marciau a ddyfarnwyd yn teimlo'n iawn.

... rydych chi'n gofyn i lawer o bobl yn ystod cymedroli, maen nhw'n gwneud hynny beth bynnag, mae bron fel eu bod yn ei farcio ddwywaith mewn ffordd.

Rydych chi'n gwybod bod y cyfan yn oddrychol. Mae'n cymryd amser. Rydych chi'n rhagweld drwy ddyfalu yn aml, mewn ffordd. Rydych chi'n mynd yn ôl ac yn ailddarllen ac yn dweud, "A ddylwn roi dau farc arall i hwnnw?"

Canfuwyd y duedd i gymharu ymatebion mewn astudiaethau eraill hefyd (e.e. Daal et al., 2016) ac roedd ein cyfranogwyr o'r farn ei bod hyd yn oed yn fwy amlwg pan fydd athrawon yn marcio eu dysgwyr eu hunain, oherwydd efallai y bydd athrawon am sicrhau bod y marciau'n adlewyrchu'r drefn restrol y maent wedi'i neilltuo i'w dysgwyr (os yn feddyliol yn unig).

Ystyriwyd bod BG yn fwy effeithlon, nid yn unig am ei bod eisoes yn mynd i'r afael â'r angen i gymharu'r ymatebion, ond hefyd am ei bod yn dileu'r angen i fynd yn ôl ac ail-wirio penderfyniadau blaenorol.

Anfantais BG yw'r angen i wneud llawer o gymariaethau sy'n gallu teimlo'n hynod anodd.

Mae hynny'n iawn, os cewch chi rifydd ar y brig sy'n dweud 500, ac yna mae'n marcio i lawr ac rydych chi'n meddwl "O dw i wedi gweithio'n galed" ac rydych chi'n edrych arno ac yn meddwl "Mae gen i 394 ar ôl", rwy'n credu y gallai hynny beri ychydig o straen.

7.1.4. Cyfleustra

Roedd gan y cyfranogwyr farn gymysg ar ba mor gyfleus fyddai asesu ymatebion gan ddefnyddio BG o'i chymharu â marcio confensiynol.

Soniodd rhai cyfranogwyr, oherwydd eu bod am sicrhau bod marciau'n gyson ar draws yr ymatebion, fod yn well ganddynt eu marcio yn ystod cyfnod byr. Nid yw'r pwysau hwn yn bodoli yn achos BG, felly gellir asesu dros gyfnodau hirach. Mae hyn yn fantais sylweddol, gan y gellir cynnal BG pryd bynnag y bydd gan yr athro amser i wneud hyn, ar yr amod bod rheolaethau'n bodoli fel nad yw athrawon sy'n llunio barn ar wahanol adegau yn ystumio canlyniadau BG.

Fodd bynnag, roedd rhai cyfranogwyr o'r farn y gallai fod yn anos asesu ymatebion yn ystod egwylliau ysgol neu amser cinio fel y maent yn ei wneud ar hyn o bryd pan fyddant yn marcio. Roeddent yn credu y byddai torri ar draws, sy'n debygol mewn amgylchedd ysgol, fwy na thebyg yn gwneud iddynt ddechrau darllen yr ymateb o'r dechrau. Nid yw hyn yn digwydd yn ystod marcio, gan y gallai athrawon ddefnyddio'r anodiadau i'w hatgoffa am ansawdd yr ymateb.

Ac mae amgylchedd yr ysgol, mae'n debyg, yn tarfu arnoch gryn dipyn rhag gallu eistedd yno am ddwy awr ar rywbeth, neu awr ar rywbeth. Felly, rwy'n meddwl byddai [BG] yn rhywbeth y byddai'r rhan fwyaf o bobl fwy na thebyg yn ei wneud y tu allan i'r ysgol yn hytrach nag yn yr ysgol, ac yna nid wyf yn gwybod a fyddent mor awyddus i wneud hynny oherwydd gall athrawon fod yn eithaf amddiffynnol o'u hamser y tu allan i ystafell ddosbarth ...

Roedd gan y cyfranogwyr safbwyntiau cymysg am fanteision defnyddio cyfrifiadur i asesu ymatebion. Roedd rhai cyfranogwyr o'r farn mai mater o arfer yn bennaf yw dewis marcio ymatebion papur. Pan ddaw athrawon i arfer â rhyngwyneb meddalwedd BG, byddent yn ei chael yn haws ac yn gyflymach asesu'r ymatebion ar y sgrin na marcio ymatebion ar bapur. Dywedodd y cyfranogwyr fod marcio ar y sgrin yn rhoi mynediad cyflymach iddynt at yr ymatebion. Yn ogystal, mae cael gliniadur neu lechen sgrin gyffwrdd yn gwneud y broses yn haws byth nag wrth ddefnyddio llygoden.

Mae hwnnw'n un rhyfedd braidd. Byddwn wedi dweud yn y gorffennol nad oeddwn yn hoffi marcio ar y sgrin nes i mi ddod i'r sesiwn feirniadu, ac yna fe'm synnwyd gymaint nad oedd ots gennyf. Roeddwn i'n meddwl na fyddwn i'n gallu, roeddwn i'n meddwl y byddwn i'n colli pethau ar sgrin na fyddwn i o reidrwydd yn eu colli pe bawn i wedi'i ysgrifennu o'm blaen, ac yn gwneud hynny'n gorfforol. Doeddwn i ddim wir yn teimlo fel hynny ...

Fodd bynnag, mae marcio ymatebion papur yn cynnwys ei fanteision ei hun. Er enghraifft, gall athrawon fynd â'r papurau i ffwrdd gyda nhw'n hawdd a marcio pryd bynnag y bo'n gyfleus. Mae'r cyfleoedd hyn yn brinnach yn achos BG gan fod angen cyfrifiadur/llechen gyda chysylltiad rhyngrwyd dibynadwy. Fodd bynnag, y fantais yw na fydd yn rhaid i athrawon boeni mwyach am ddiogelwch yr ymatebion.

Rwy'n credu eich bod wedi'ch rhwymo'n fawr wedyn wrth ... mae'n rhaid i chi eistedd wrth eich desg. Ond gyda chopi caled, fe allech ei wneud ar y trê, gallech ei wneud yn unrhyw le.

7.1.5. Baich gweinyddol

Awgrymodd y cyfranogwyr y gallai BG leddfu'r gwaith gweinyddol sy'n ofynnol ar hyn o bryd mewn perthynas ag ymatebion yr NEA.

Pan fydd dysgwyr yn cwblhau eu hymatebion, mae athrawon yn casglu ffurflen wedi'i llofnodi gan ddysgwyr sy'n cadarnhau mai eu gwaith nhw yw'r gwaith y maent wedi'i gyflwyno. Mae athrawon yn storio'r papurau wedi'u cwblhau mewn lle diogel nes bod y marciau wedi'u cadarnhau. Oherwydd y gall y dysgwyr ymgymryd â'r tasgau ar

unrhyw adeg yn ystod y cwrs, gall y storio gymryd misoedd. Soniodd rhai athrawon y gall hyn achosi rhywfaint o bryder gan y byddent yn poeni am gadw'r papurau'n ddiogel.

Pan fydd sampl o bapurau'n cael eu dewis ar gyfer y safoni allanol, mae angen i athrawon eu hanfon drwy'r post i CBAC. Soniodd rhai cyfranogwyr y gallai hyn greu pwysau ychwanegol os yw'r dyddiad cau ar gyfer cyflwyno tuag adeg gŵyl banc.

Teimlai'r cyfranogwyr y byddai BG yn gwneud y broses weinyddol hon yn symlach oherwydd byddai'r papurau'n cael eu lanlwytho i safle penodol.

Rwy'n credu bod prosesau marcio papur [...] mor llafurddwys ac mae postio a phryderu am sgriptiau'n mynd ar goll a'r mathau hynny o bethau. Ond rwy'n meddwl os yw popeth wedi'i lanlwytho i system mae llawer mwy o dawelwch meddwl ei fod yn cael ei gadw'n ddiogel.

Fodd bynnag, mae'r fantais hon yn rhagdybio bod dysgwyr yn teipio eu hymatebion. Rhaid sganio ymatebion llawysgrifen, a byddai hyn yn gwrthbwysu'r fantais oherwydd teimlwyd y gallai sganio gymryd hyd yn oed mwy o amser nag sydd ei angen ar hyn o bryd ar gyfer y gwaith gweinyddol.

Serch hynny, dylem nodi nad manteision BG yn unig mo'r cyfleustra gweinyddol a marcio, ac y gellid cyflawni hyn gyda marcio confensiynol ar y sgrin hefyd.

7.2. Yr effaith ar ddatblygiad proffesiynol

Soniodd y cyfranogwyr fod cael adborth ar ansawdd yr ymatebion y mae eu dysgwyr yn eu cynhyrchu yn ffordd bwysig o wella eu haddysgu. Daw'r adborth a gânt o gymedroli mewnol neu allanol. Roedd rhai cyfranogwyr o'r farn bod adroddiadau cymedrolwyr yn ddefnyddiol ac yn poeni na fyddai athrawon yn cael yr adborth hwnnw mwyach pe bai BG yn cael ei gweithredu. Ar y llaw arall, soniodd cyfranogwyr fod y cymedroli'n bodoli er mwyn sicrhau cysondeb marcio. Felly, dim ond rhan gyfyngedig o'r agweddau a allai fod yn berthnasol i'w datblygiad proffesiynol y mae unrhyw adborth sy'n deillio o gymedroli yn ei chwmpasu.

At hynny, credai rhai cyfranogwyr nad yw'r wybodaeth a geir yn ystod cymedroli bob amser yn ddefnyddiol nac yn ddibynadwy. Roeddent yn teimlo bod meddwl grŵp yn effeithio ar gymedroli mewnol, tra nad oedd adroddiadau cymedroli allanol bob amser yn benodol, a oedd yn cyfyngu ar ba mor ddefnyddiol y gallent fod. Soniodd y cyfranogwyr, er enghraifft, y byddai'r adroddiad cymedroli allanol yn aml yn cynnwys ymadroddion cyffredinol yn unig, megis bod y 'marcio'n iawn' neu fod y marcio'n 'rhy lym neu'n rhy drugarog'. Cyfyngir ymhellach ar ddefnyddioldeb yr adroddiad hwn os yw'r un cymedrolwr yn adolygu system farcio ysgolion dros y blynyddoedd.

Wel, nid yw'n wych ar hyn o bryd, oherwydd ar hyn o bryd rydych yn cael adroddiad cymedrolwr, ac mae'n dweud wrthy ch a oeddent yn cytuno â'ch marcio, os oeddent yn meddwl eich bod yn anghywir, ond nid yw'n rhoi unrhyw fanylion i chi mewn gwirionedd. Felly, efallai y byddant yn dweud wrthy ch y buoch yn rhy hael, ond nid yw'n dweud wrthy ch pa mor hael neu a oedd yn y pen isaf neu'r pen uchaf. Felly, mae'n dipyn o enigma. Rydych chi'n anfon y gwaith hwn i ffwrdd, a rhywun yn dweud "Ydy, mae'n iawn," ac yna rydych chi'n ailadrodd y broses y flwyddyn ganlynol.

I'r gwrthwyneb, roedd y cyfranogwyr o'r farn bod beirniadu ymatebion ysgolion eraill yn gyfle datblygiad proffesiynol gwych gan ei fod yn caniatáu i athrawon weld sut mae ysgolion eraill yn mynd at dasg yr NEA a chael gwell dealltwriaeth o sut mae safonau'n cymharu ar draws canolfannau.

Roedd yn agoriad llygad o sut y gallem ... sut gallwn i osod y gwaith cwrs ar gyfer fy myfyrwyr fy hun a ... wyddoch chi, fel rhannu arfer da mewn gwirionedd.

O weld rhai o'r [ymatebion], yn enwedig y rhai Jack the Ripper yr wyf yn eu haddysgu, rwy'n teimlo, "O iawn, rwy'n bendant ar y trywydd cywir. Rwy'n dweud y pethau cywir wrth y plant."

... yn ein hysgol ni rydyn ni mewn tipyn bach o swigen. Felly, er fy mod yn credu y gallai fy nhasgau NEA fod yn dda iawn, gallwn edrych ar rywun sydd mewn ysgol arall yng Nghaerdydd a gallai fy un i fod yn weddol. Felly, er y gallwn raddio'n fewnol, nid wyf mewn gwirionedd yn graddio i'r darlun cenedlaethol. Ond rwy'n credu y byddai [BG] yn help mawr.

8. Gweithredu BG fel dewis yn lle marcio confensiynol

Teimlai'r cyfranogwyr y byddai BG yn hawdd ei chynnal gan athrawon profiadol a dibrofiad. Roedd mwyafrif helaeth o'r cyfranogwyr o'r farn bod BG yn fwy addas ar gyfer asesu ymatebion rhan (a) NEA na marcio. Pe na fyddai BG yn cymryd mwy neu lawer mwy o amser na marcio confensiynol, dywedodd y cyfranogwyr hyn y byddai'n well ganddynt asesu gan ddefnyddio BG.

Mae'r adran hon yn archwilio'r agweddau y soniodd y cyfranogwyr y byddai angen eu hystyried cyn y gellid gweithredu BG.

8.1. Cynnwys a hysbysu athrawon

Er bod BG fel petai'n arf gwych ar gyfer asesu, rhybuddiodd y cyfranogwyr y byddai athrawon yn gwrthod ei gweithredu pe byddent yn teimlo ei bod yn cael ei gorfodi arnynt. Soniodd y cyfranogwyr hefyd y gallai athrawon fod yn amharod i newid rhywbeth y maent eisoes yn ei wybod, ac yn gyfarwydd ag ef, am rywbeth newydd. Roedd ambell gyfranogwr hefyd yn dweud mai'r gwrthwynebiad i newid oedd y rheswm y byddai'n well ganddynt farcio confensiynol o hyd.

Byddwn i'n mynd am y traddodiadol ond wedyn rydw i ar ddiwedd fy ngyrfa felly, dyna dw i wedi arfer ag ef o hyd. Pe bawn i ar ddechrau fy ngyrfa, mae'n debyg y byddwn yn mynd am hynny. Byddwn i'n mynd am y ffordd arall.

Ond, o'm safbwynt i, rwy'n defnyddio [marcio] traddodiadol ers dros 20 mlynedd ac rwy'n credu bod gen i hyder naturiol yn marcio felly. Felly, byddwn yn tueddu i ddweud y traddodiadol.

Credai'r cyfranogwyr y dylid hysbysu athrawon am BG, a'u cynnwys wedyn mewn llunio barn, fel y gallent ddod i arfer ag ef (e.e. fel offeryn datblygu ffurfiannol neu broffesiynol). Credai'r cyfranogwyr y byddai'r rhan fwyaf o athrawon yn ffafrio BG ar ôl iddynt geisio ei deall.

Awgrymodd y cyfranogwyr y byddai rhai o'u cydweithwyr yn amharod i symud o farcio ar bapur i farcio ar y sgrin. Rhoddodd rhai cyfranogwyr enghreifftiau o farcwyr yr oeddent yn gwybod eu bod wedi gwrthod marcio pan ofynnodd y corff dyfarnu iddynt wneud hynny ar y sgrin. Byddai cael y cyflwyniad ysgafn hwn i BG yn helpu athrawon o'r fath i oresgyn eu hamharodrwydd neu hyd yn oed eu hofn ynghylch gorfod gweithio gyda thechnoleg.

Un agwedd arall yr awgrymodd cyfranogwyr y byddai'n gwneud athrawon yn amharod i dderbyn BG oedd yr ansicrwydd ynghylch y canlyniadau terfynol, a'r pryder ynghylch a fyddai eu dysgwyr yn cael y radd y maent yn ei haeddu. Soniodd y cyfranogwyr fod y system farcio bresennol yn rhoi mwy o sicrwydd i athrawon am hyn.

Rwy'n gwybod i raddau helaeth iawn, bron 100%, mai'r marc a roddaf iddynt yw'r marc terfynol y byddant yn ei gael. Felly dyna'r prif beth. Yr anhysbys yw'r peth.

Rwy'n credu y bydd pobl yn meddwl, "Wel, sut ydw i'n gwybod bod y bobl sy'n llunio'r farn am ein canolfan yn ei weld fel y byddem ni?" Rwy'n credu y byddai pobl yn ansicr iawn.

Byddai'n frawychus fel athro i feddwl, "Wn i ddim ble fydd y marc hwnnw nawr pan ddaw'n ôl," oherwydd, rydych chi'n gwybod, rwy'n dibynnu ar, "Iawn, mae hyn y ffordd hon," gwnewch beth bynnag, am eu gwaith cwrs.

Roedd fel petai'r pryder am y marciau terfynol hefyd yn deillio o'r ofn y byddai athrawon yn beirniadu ymatebion gan ysgolion sydd â lefelau perfformiad gwahanol o gymharu â'u hysgol eu hunain. Roedd athrawon o ysgolion uwch eu cyrhaeddiad yn pryderu y byddai eu dysgwyr yn cael eu marcio i lawr, gan na fyddai beirniaid yn ystyried agweddau ychwanegol sy'n ychwanegu gwerth at ymateb. Ar y llaw arall, roedd cyfranogwyr o ysgolion is eu cyrhaeddiad yn poeni y byddai gwaith eu dysgwyr yn edrych yn wael o'i gymharu ag ysgolion â dysgwyr yn dod o gymdogaethau mwy breintiedig.

Rwy'n teimlo ychydig yn anesmwyth bod rhywun arall yn marcio fy ngwaith cwrs, os yw eu gwaith cwrs yn ofnadwy; bydd hynny'n un anfantais.

... oherwydd eu bod yn dod o ardaloedd da iawn, gall y disgyblion hynny yn naturiol ysgrifennu'n well na rhai o'r disgyblion rwy'n eu haddysgu [sy'n] ei chael hi'n anodd darllen. Felly os ydych chi'n mynd i gael ateb gan un o'r ysgolion mwy clodfawr ac yn mynd i gael ateb gan un o'm disgyblion i [...] felly bydd fy nisgybl bob amser yn cael ei roi i lawr o'i gymharu â hynny a byddwn yn teimlo bod hynny'n annheg iawn oherwydd er y gallent fod yn cyrraedd y sgiliau, efallai na fydd eu rhuglder cystal a byddwn yn teimlo y byddai hynny'n dipyn o anfantais iddynt.

Roedd y cyfranogwyr o'r farn y gallai fod gan athrawon bryderon tebyg mewn perthynas â dysgwyr gradd C ffiniol y teimlent y gellid eu 'gwthio' ar hyn o bryd i gael C.

Credwn y gellid mynd i'r afael â'r pryderon hyn drwy roi gwybod yn briodol i athrawon sut y gellir cyrraedd canlyniadau wrth ddefnyddio BG. Gallai helpu hefyd os pwysleisir mesurau diogelu BG i atal penderfyniadau mympwyol (fel *misfit*), os bydd mwy o athrawon yn edrych ar yr un ymateb, os bydd yn bosibl cael proses ar wahân o bennu graddau a ffiniau graddau.

8.2. Canllawiau BG i athrawon a dysgwyr

Nododd y cyfranogwyr y byddai angen arweiniad clir ar athrawon ar sut i ddewis yr ymateb gwell. Roeddent o'r farn bod y cynllun marcio presennol yn cynnwys yr agweddau perthnasol, felly, awgrymwyd y gellid ei ddefnyddio fel man cychwyn wrth ddrafftio'r canllawiau. Fodd bynnag, ystyriwyd y dylai'r canllawiau newydd fod yn fwy

cyfannol a chynnig mwy o 'dragwyddol heol' a sicrhau bod dysgwyr yn cael cyfle i ysgrifennu ymatebion mwy creadigol.

Felly, mae angen ichi gael rhyw fath o bwyntiau cyfeirio er mwyn i'r arholwr edrych i weld beth i chwilio amdano. Mae angen i hyn fod ar ffurf sgaffaldiau heb gyfyngu ar greadigrwydd.

Awgrymodd y cyfranogwyr hefyd y dylid disgrifio'r agweddau a gynhwysir yn y canllawiau ar lefel fwy cyffredinol ac y dylent osgoi gormod o fanylion. Nodwyd cynllun marcio Safon Uwch Hanes ar gyfer NEA fel enghraifft o ddull gweithredu o'r fath. Yn eu barn hwy, dylai'r canllawiau gynnwys bandiau o ddisgrifyddion, er enghraifft, yr hyn sy'n dda, yn dda iawn ac yn rhagorol. Roeddent hefyd yn teimlo mai dyma sut y defnyddir y cynllun marcio presennol beth bynnag.

Nid yw'r cynlluniau marcio mewn gwirionedd yn gymaint â hynny o ganllaw, ac weithiau mae'n fwy o reddf neu weddu i'r lefel honno. Hyd yn oed wrth farcio. Wyddoch chi, rydych chi'n clywed pethau mewn cynadleddau lle mae llawer o anghydfod o ran goddefgarwch gydag ymatebion oherwydd bod y cynlluniau marcio yn amleiriog iawn. A gallai fod, dim ond newid un gair sydd wedyn yn eich troi o gael lefel, Band 5 i Fand 6. Felly, rwy'n credu pe baem yn dilyn y llwybr barn gymharol, rwy'n credu y byddai angen i'r cynlluniau marcio newid fel eu bod ychydig yn fwy cyfannol ...

Ar wahân i ganllawiau i athrawon, teimlai'r cyfranogwyr y byddai angen fframwaith a fyddai'n amlinellu'r disgwyladau neu'r meini prawf llwyddiant i ddysgwyr. Roedd y cyfranogwyr o'r farn y gallai defnyddio canllawiau i athrawon fod yn anodd i ddysgwyr ei ddeall, fel sy'n wir am y cynllun marcio presennol. Felly, dewisasant gael dogfen ar wahân mewn iaith glir i ddysgwyr.

Ond, wyddoch chi, allwn ni ddim helpu ond mynd, "Foneddigion a boneddigesau, dyma'r cynllun marcio, beth ydych chi'n ei ddeall wrth hynny?" A dydyn nhw ddim, ac yna'r frwydr hon i fyny'r allt.

Gellir defnyddio BG i greu graddfa o enghreifftiau wedi'u graddnodi y gellir eu defnyddio wedyn wrth asesu ymatebion (McGrane, J. A, 2018). Gallai hyn fod yn ffordd arall o gynnal cysondeb, er y byddai wedyn yn ddull gwahanol o asesu ymatebion na'r un a drafodir yn yr astudiaeth hon.

8.3. Hyfforddiant

Teimlai'r cyfranogwyr fod gweithio gyda meddalwedd BG yn hawdd a byddai rhai ymarferion ymarfer byr yn ddigon i athrawon ddod i arfer â hi. Awgrymodd rhai y byddai gweithgaredd byr, fel yr un a gynhaliwyd ganddynt ar ddechrau'r gweithdy,

pan oeddent yn cymharu darllenadwyaeth ysgrifennu, yn ddigon i athrawon ddysgu'r agweddau technegol ar lywio'r feddalwedd a llunio barn.

Credai'r cyfranogwyr y byddai angen mwy o hyfforddiant i addysgu beirniaid sut i wneud y penderfyniadau'n gywir. Fodd bynnag, roeddent o'r farn bod beirniadu'n gymharol hawdd, ac nad oedd angen hyfforddiant sylweddol, neu o leiaf nad oedd angen hyfforddiant sy'n fwy beichus na'r hyfforddiant y mae athrawon eisoes yn ei gael i fod yn farcwyr.

Yn rhan o'u hyfforddiant, byddai angen i athrawon ymgyswrtu â sut mae BG yn gweithio (gan gynnwys hap-drefn ymatebion yn y parau pan nad ydynt wedi'u cadwyno) ac ymgyswrtu â'r canllawiau beirniadu. Yna dylai'r rhan ddamcaniaethol hon gael ei dilyn gan sesiwn o BG o dan oruchwyliaeth, neu mewn grwpiau, a fyddai'n cynnig y cyfle i gael adborth ar y farn a luniwyd. Yna ar ôl y cam hwn dylid cael sesiwn o BG gan ddefnyddio ymatebion gyda gwerthoedd hysbys.

Roedd rhai cyfranogwyr o'r farn y gellid cynnal hyfforddiant BG mewn ysgolion, tra bod eraill yn awgrymu y gallai ddigwydd ar-lein. Awgrymwyd hefyd y byddai cynnwys athrawon fel hyfforddwyr yn ddymunol.

Byddwn yn dweud y byddai athrawon yn rhoi hyfforddiant i athrawon eraill hefyd yn dda oherwydd ein bod yn tueddu i ymddiried yn ein gilydd fel gweithwyr proffesiynol ...

Teimlai'r cyfranogwyr y byddai rhoi gwybod i feirniaid am ansawdd eu penderfyniadau (e.e. pa mor bell i ffwrdd o'r consensws oedden nhw, neu amlygu ymatebion a oedd yn anodd eu beirniadu) yn helpu athrawon ymhellach i raddnodi eu sgiliau beirniadu. Gellid ategu'r hyfforddiant gyda mewnwelediadau ychwanegol i farn effeithiol pan fyddai'r wybodaeth hon ar gael.

Oherwydd i mi edrych dros fy nodyn hefyd ac roeddwn i'n meddwl ... ac ar ôl ... do, oherwydd sawl gwaith pa un bynnag a ddewisais ar gyfer B ymddangosodd yn A. Digwyddodd hynny sawl gwaith, ac roeddwn i'n meddwl, "O, pam mae hynny'n digwydd? A yw B yn well nag A?" Neu i'r gwrthwyneb.

8.4. Hyd yr ymatebion

Roedd cyfranogwyr o'r farn bod beirniadu ymatebion hir yn anoddach na rhai byrrach. Roeddent yn teimlo bod atebion hir yn ei gwneud yn fwy heriol cadw'r wybodaeth ac y gallai bwrw golwg cyflym ddod yn dechneg lai effeithlon oherwydd gellid colli agweddau pwysig.

Mae ymatebion byr hefyd yn haws penderfynu yn eu cylch oherwydd eu bod yn creu teimlad o gynnydd cyflymach. Ar y llaw arall, gellid beirniadu ymatebion eithriadol o hir yn fwy llym gan y gallant greu tuedd negyddol yn eu herbyn.

Fel y soniwyd yn gynharach, mae'r fanyleb yn nodi bod ymatebion rhan (a) yn cynnwys tua 1,000 o eiriau. Ond argymhelliad yn unig yw hwn ac nid yw'n orfodol. Felly, mae llawer o'r ymatebion yn llawer hwy na'r hyd a argymhellir. Credai'r cyfranogwyr y dylid gorfodi uchafswm neu ystod uchaf o eiriau, a chredai'r rhan fwyaf ohonynt y byddai ystod o 1,000-1,500 o eiriau yn dderbyniol ar gyfer rhan (a). Byddai hyn yn cynyddu'r gallu i reoli'r broses BG heb fawr o risg i ddilysrwydd yr asesiad.

8.5. Mathau o ymateb

Roedd gan gyfranogwyr farn gymysg ynghylch a allai beirniadu ddod yn anoddach, neu'n llai anodd, gan ddibynnu ar yr hyn a asesir. Teimlai rhai cyfranogwyr fod ymatebion rhan (a) yn syml i'w beirniadu ond byddai rhan (b), sy'n draethawd dehongli, yn fwy heriol yn wybyddol. Ar y llaw arall, teimlai rhai cyfranogwyr y byddai ymatebion rhan (b) yn haws i'w cymharu oherwydd eu bod yn fwy cymhleth, sy'n golygu y byddai'n haws gwahaniaethu rhwng ymatebion.

Yn yr un modd, roedd cyfranogwyr o'r farn y gallai BG fod yn briodol i farcio NEA Safon Uwch Hanes gan fod y dull asesu mewn Safon Uwch yn fwy cyfannol nag ar gyfer TGAU. Yr unig anfantais i'r awgrym hwn yw bod ymatebion Safon Uwch yn hir (4,000 o eiriau) a'u bod wedi'u hysgrifennu ar bynciau amrywiol, a fyddai'n gwneud cymariaethau'n anoddach.

Disgwylir y gallai ymatebion o wahanol gymhlethdodau ddylanwadu ar rwyddineb beirniadu. Fodd bynnag, nid yw'n glir sut y byddai hyn yn dod i'r golwg ac a fyddai'r galw'n wahanol i farcio.

8.6. Yr amser sydd ei angen

Fel y soniwyd yn gynharach, roedd barn cyfranogwyr am BG fel offeryn marcio yn tybio na fyddai angen mwy, neu lawer mwy, o amser na marcio. O dan y dybiaeth hon, roedd y rhan fwyaf o'r cyfranogwyr o'r farn bod BG yn well na marcio traddodiadol. Un cwestiwn pwysig ar gyfer gweithredu BG yn llwyddiannus felly yw gofyn a yw'r dybiaeth amser tebyg yn gywir. Mae dau newidyn yn cael effaith bwysig ar yr amser sydd ei angen i gwblhau BG: amser beirniadu cyfartalog a nifer y barnau sydd eu hangen.

Yr amser canolrif ar gyfer llunio barn yng Ngrŵp NF oedd tuag wyth munud. Os tybiwn y byddai'r ymatebion ar gyfer y dasg rhan (a) yn cael eu capio ar 1,500 o

eiriau ac y byddant i gyd yn cael eu teipio, gallem obeithio y dylai barn gymryd tua phum munud ar gyfartaledd.

Mae Tabl 4 yn dangos yr amser bras sydd ei angen ar gyfer tasgau amrywiol o dan y trefniadau marcio presennol, yn seiliedig ar ddsbarth o 22 o ddysgwyr. Dywedodd y rhan fwyaf o'r cyfranogwyr y byddent yn nodi tri i bedwar ymateb yr awr. Os cymerwn mai hwn yw'r amser marcio cyfartalog tybiedig, mae'n golygu y byddai angen tua chwe awr ar athro i farcio 22 o ymatebion. Soniodd y cyfranogwyr hefyd y gallai gymryd hyd at ddiwrnod i gwblhau'r gwaith papur ar gyfer un dosbarth, uwchlwytho'r marciau, ac anfon y papurau a ddewiswyd i'w cymedroli'n allanol. Gallai gymryd hanner diwrnod arall i ddiwrnod i gymedroli'r marcio'n fewnol. Felly, cyfanswm yr amser sydd ei angen ar hyn o bryd yw tua 18-22 awr.

Tabl 4: Brasamcan o'r amser a dreulir ar asesu traddodiadol yn seiliedig ar 22 o ddysgwyr

Tasg	Marcio ymatebion	Tasgau gweinyddol	Tasgau cymedroli	Cyfanswm
Amser (oriau)	6	8	4-8	18-22

Ni fyddai angen cymedroli mewnol nac allanol ar BG ac, ar yr amod bod yr ymatebion yn cael eu teipio, byddai BG yn mynnu amser ansylweddol ar gyfer y tasgau gweinyddol sy'n gysylltiedig â phrosesu'r atebion. O ystyried hyn, byddem yn tybio, er mwyn sicrhau'r un effeithlonrwydd amser, na ddylai'r beirniadu gwirioneddol gymryd mwy na thua 17 awr ar gyfer dosbarth o 22 o ddysgwyr (gan ganiatáu 1-5 awr ar gyfer tasgau gweinyddol ac adolygu). Mae hyn yn golygu, er mwyn sicrhau effeithlonrwydd amser tebyg, y byddai angen i ni gyfyngu ar nifer y barnau fesul ymateb i ryw 9-10. Mae hyn yn debyg i nifer y barnau a gynhaliwyd yn yr astudiaeth hon.

Fel y soniasom yn gynharach, y dibynadwyedd yng Ngrŵp NF oedd 0.74 (rydym yn canolbwyntio ar y grŵp hwn am ein bod yn ystyried mai peidio â darparu adborth fydd y dull tebygol o ymdrin â BG pan gaiff ei defnyddio mewn cyd-destun asesu go iawn, a hefyd am fod ganddynt ddibynadwyedd is a ddylai sicrhau dull mwy ceidwadol o ymdrin â'r drafodaeth sy'n dilyn). Nododd Bramley (2015) fod ymaddasolrwydd yn chwyddo'r cyfernodau dibynadwyedd ac awgrymodd Bramley a Vitello (2019) fod angen 16 barn fesul gwrthrych er mwyn sicrhau dibynadwyedd da⁸. Fel y nodwyd gennym yn gynharach, ni ddylai chwyddiant ddigwydd gyda'r feddalwedd a ddefnyddiwyd gennym, gan fod ei algorithm wedi'i optimeiddio i

⁸ Mewn astudiaeth efelychu arall a gynhaliwyd gan Cromptvoets, E. A. et al. (2020), fe ganfuwyd bod angen 20 cymhariaeth ar gyfartaledd i gyrraedd dibynadwyedd o .80.

leihau'r chwyddiant. Fel y dengys Kimbell (2021) cytunodd Bramley y byddai'r newidiadau a wnaed gan RM i'w feddalwedd yn golygu y byddai 15 cylch beirniadu (h.y. 7.5 barn fesul sgript) yn dileu risg chwyddiant dibynadwyedd. O ystyried hyn, byddem yn tybio nad oedd chwyddiant yn effeithio ar y dibynadwyedd 0.74 a gyflawnwyd gan Grŵp NF. Os cymerwn hwn fel gwerth sylfaenol, a fyddai'n dderbyniol pe bai marcio confensiynol yn cael ei ddisodli?

Mae cwestiynau o fath traethawd yn canfod y lefelau isaf o ddibynadwyedd marcio confensiynol oherwydd yn wahanol, er enghraifft, i brawf mathemategol, mae'n anodd iawn penderfynu'n union sut i ddyrannu pob marc a gynhwysir yn y cynllun (Kellaghan et al. 2019 t. 167). Fel y soniwyd yn gynharach, mae astudiaeth a gynhaliwyd gan Rhead et al. (2018, t.4) wedi dangos bod tebygolrwydd cael 'gradd ddiffiniol' ar gyfer arholiadau yn Lloegr ar ei isaf ar gyfer ymatebion estynedig ar gyfer iaith a llenyddiaeth Saesneg (0.52) a hanes oedd â'r ail farcio cysondeb isaf. Canfu Holmes et al. (2018, t.17) fod dibynadwyedd marcio cwestiynau ysgrifennu arholiad Hanes UG sy'n seiliedig ar ffynhonnell o dri phapur gwahanol yn amrywio o 0.52 i 0.62.

Ni wyddom a yw dibynadwyedd y marciau confensiynol ar gyfer yr ymatebion hyn a archwiliwyd yn yr ymchwil hon yn debyg i'r hyn y mae eraill wedi'i ganfod. Ar y naill law, gallwn dybio y gallai fod yn is, o ystyried bod yr ymatebion yn hwy; ar y llaw arall, gallai fod yn uwch os ystyriwn y cymedroli allanol. Os tybiwn fod dibynadwyedd asesiad NEA TGAU Hanes tua 0.6, yna mae'n debyg y byddai 10 barn fesul ymateb yn cynnig trefn restrol sydd o leiaf mor ddibynadwy â marcio confensiynol. Fodd bynnag, byddai angen ymchwil ychwanegol i ganfod hyn cyn y gellir dod i gasgliad mwy pendant am yr agwedd hon ar ddichonoldeb BG.

Yn gyffredinol, credwn, ar gyfer ysgrifennu estynedig, ei bod yn debygol bod dibynadwyedd BG o leiafystal â'r marcio confensiynol, gan y derbynnir yn gyffredinol fod aseswyr yn well am farnau cymharol na rhai absoliwt. Mewn astudiaeth, gofynnodd Gill a Bramley (2013) i grŵp o athrawon hanes bennu graddau (A, B neu C) cyfres o bapurau arholiad Hanes (Safon Uwch mae'n debyg) gan ddefnyddio barnau cymharol ac absoliwt. Roedd y rhan fwyaf o sgriptiau yn yr astudiaeth hon o amgylch y ffin A/B a oedd yn gwneud y penderfyniadau'n anoddach (roedd y sgriptiau'n amrywio o 57 marc hyd at 75 ac roedd y ffin gradd A/B yn 68 marc tra bod y ffin B/C yn 61 marc). Er bod gan feirniaid lefelau tebyg o hyder am eu cywirdeb wrth ddefnyddio'r ddau ddull, roeddent yn dangos llai o amrywioldeb ac roeddent yn llawer mwy cywir wrth ddefnyddio barn gymharol (neilltuwyd y radd wreiddiol i 66.1% o bapurau o ganlyniad i farn gymharol o gymharu â 41.7% wrth lunio barn absoliwt).

Os yw dibynadwyedd marcio confensiynol yn uwch na'r hyn y gellir ei gyflawni ar ôl 10 barn fesul sgript, un ffordd bosibl o ddelio â hyn fyddai profi a fyddai arwain y

detholiad cychwynnol o barau drwy ddarparu gwybodaeth ychwanegol am werth yr ymatebion yn cynyddu dibynadwyedd BG gan gadw nifer y barnau'n isel (gweler mwy Bramley, 2015). Fel yr esbonia Bramley, mae goramcangyfrif mynegeion dibynadwyedd yn ystod BGA yn digwydd oherwydd nad yw gwerthoedd y gwrthrychau'n hysbys ymlaen llaw. Felly, amcangyfrifir pob paramedr ar yr un pryd wrth fwydo'r data o'r beirniadu i'r model. Mae hyn yn creu 'gwahaniad annilys' o'r gwrthrychau. Bydd cyflwyno rhywfaint o wybodaeth am werth y gwrthrychau yn lleihau risg gorchwyddiant. Yn ein hachos ni, gallai hyn fod yn rhai amcangyfrifon a ddarparwyd gan athrawon yn seiliedig ar yr wybodaeth sydd ganddynt am waith y dysgwyr. Fodd bynnag, mae angen ymchwilio ymhellach i'r syniad hwn er mwyn gwirio nad yw'r gwerthoedd cychwynnol yn wyrdueddol am wahanol resymau, a allai yn ei dro greu problemau gyda cham addasedig y broses.

9. Cyfyngiadau astudio

Ffocws yr astudiaeth oedd y dadansoddiad ansoddol o'r broses feirniadu a barn athrawon am y dull. Mae'r astudiaeth yn seiliedig ar sampl cyfleustra a sicrhaodd gynrychiolaeth resymol o athrawon. Ni allwn gadarnhau a ellid cymhwyso barn y cyfranogwyr am BG i'r boblogaeth athrawon gyfan. Er hynny, mae'r rhan fwyaf o'r farn a fynegwyd ynghylch BG yn cyd-fynd â'r hyn y mae astudiaethau sy'n casglu adborth gan feirniaid wedi'i ganfod (e.e. Pollitt 2012), er mai ychydig o astudiaethau o'r fath sydd. Felly, byddai ymchwil ychwanegol i gadarnhau'r canfyddiadau a'r casgliadau a wnaed yn yr astudiaeth gyfredol yn cael ei hargymell.

Ni chynlluniwyd yr astudiaeth gyfredol i gasglu data ar gyfer dadansoddiad ystadegol cadarn. Felly, dylid ystyried canfyddiadau ystadegol yr astudiaeth hon yn feirniadol. Er enghraifft, mae'n bosibl bod y gronfa fach o sgriptiau yn gwneud i feirniaid weld rhai ohonynt fwy nag unwaith mewn cyfnod cymharol fyr. Gallai hyn fod wedi effeithio ar y dybiaeth annibyniaeth a fyddai, yn ei thro, yn ei gwneud yn anos dehongli'r cyfernodau ystadegol a'r mynegeion. Gallai'r defnydd o sylwadau yn un o'r grwpiau fod wedi gwaethygu hyn ymhellach, gan fod sylwadau a wnaed yn flaenorol wedi'u defnyddio pan feirniadwyd y sgript eto. Gallai gwneud sylwadau hefyd fod wedi effeithio ar natur gyfannol y barnau. Yn olaf, gan dybio y gallai ymatebion hwy fod yn anos eu beirniadu'n gyson, mae'n debyg bod angen mwy o farnau fesul sgript i arsylwi ar ddeinameg mynegeion dibynadwyedd.

10. Casgliadau ac ymchwil bellach

Nod allweddol yr astudiaeth hon oedd archwilio barn athrawon am ddefnyddio BG i asesu ymatebion ysgrifenedig estynedig rhan (a) NEA TGAU Hanes a nodi'r agweddau y maent yn eu hystyried wrth feirniadu.

Credai'r cyfranogwyr fod gan BG sawl mantais dros farcio confensiynol. Y pedair mantais amlwg oedd:

- rhwyddineb cymharol llunio barn gymharol;
- lliniaru tueddiadau penodol mewn asesiadau a gynhelir gan athro/athrawes dysgwr;
- y potensial ar gyfer datblygiad proffesiynol; a'r
- potensial i leihau'r ymddygiadau annymunol sy'n gysylltiedig â thargedu'r cynllun marcio.

Byddai'r dull hefyd yn darparu mwy o wybodaeth asesu (e.e. amcangyfrifon o ddibynadwyedd yn seiliedig ar y cofnod cyfan), nad yw ar gael ar hyn o bryd ar gyfer NEA a farciwyd yn fewnol. Yn gyffredinol, byddai cyfranogwyr yn hapus i ddefnyddio BG, ar yr amod nad yw'n cymryd mwy neu lawer mwy o amser na marcio confensiynol.

Dengys yr astudiaeth hon fod BG yn debygol o fod yn ddewis dilys yn lle marcio a chymedroli ymatebion rhan (a) NEA TGAU Hanes. Mewn cytundeb â Whitehouse (2013)⁹, dengys y canfyddiadau fod athrawon, at ei gilydd, yn cymhwyso'r un meini prawf ag y byddent yn eu cymhwyso wrth farcio. Er i Whitehouse (2013) ganfod bod beirniaid wedi hepgor rhai o feini prawf y cynllun marcio, ni welsom dystiolaeth o hyn, gan fod yr holl agweddau a gynhwyswyd mewn amcanion asesu wedi'u crybwyll. Fodd bynnag, yn unol â phwysiad y cynllun marcio, soniwyd llai am yr agweddau a oedd yn cario llai o farciau (e.e. ansawdd cyfathrebu ysgrifenedig) a allai awgrymu eu bod yn llai pwysig wrth feirniadu.

Nid oedd cyfranogwyr yn gwbl ymwybodol o'r broses o wneud penderfyniadau. Awgryma'r canfyddiadau y gallai rhai cyfranogwyr fod wedi cymhwyso rheolau mwy darbodus wrth wneud penderfyniadau (e.e. y cyfranogwr a lwyddodd i feirniadu'n llawer cyflymach na'r gweddill) tra bod eraill efallai'n dueddol o ddefnyddio dulliau mwy cymhleth, dadansoddol (e.e. beirniaid yn pennu marc i bob sgript). Mae'n debygol na fyddai beirniaid yn ystyried pob agwedd ar y cynllun marcio ym mhob barn. Gall hyn ddigwydd yn arbennig pan fydd beirniaid yn credu y gallant wneud y penderfyniad ar sail bwrw golwg cyflym neu ddarllen y sgriptiau'n anghyflawn. Mewn achosion felly, mae'n bosibl bod beirniaid yn cymharu'r ddwy sgript drwy edrych ar un neu ddwy agwedd sydd o fwy o bwys. Os oes gwahaniaeth clir rhwng y ddwy sgript, byddent yn penderfynu heb ddarllen ymhellach neu'n fanylach, er nad yw hyn yn wir bob amser o reidwydd. Os nad oes gwahaniaeth clir, byddai beirniaid yn ystyried agweddau eraill, wrth hefyd ddarllen y sgriptiau'n fwy trylwyr. Gallai'r dull hwn esbonio pam mai dim ond rhai agweddau a grybwyllwyd gan feirniaid yn eu sylwadau neu yn ystod sesiynau meddwl yn uchel a pham y soniwyd yn amlach am yr agweddau pwysicach.

Yn hyn o beth, byddai'n ddefnyddiol archwilio sut y gall ystyriaeth ddetholus o feini prawf asesu effeithio ar ddilysrwydd y barnau, ac a ellir dyfeisio set o reolau neu

⁹ Roedd gan Daal et al. (2016) ganfyddiadau tebyg mewn lleoliad prifysgol.

hewristeg syml ar gyfer cymharu sgriptiau a fyddai'n cynyddu dilysrwydd ac effeithlonrwydd.

Yn debyg i Whitehouse (2013), gwelsom fod cyfranogwyr yn dibynnu ar agweddau nad oeddent wedi'u cynnwys yn y cynllun marcio. Yn ein hastudiaeth, soniodd y cyfranogwyr am yr arddull ysgrifennu fecanistig fel y'i gelwir. Defnyddiwyd yr agwedd hon i dorri'r ddadl, ond roedd ganddi rôl hefyd o ran creu argraff gyffredinol o sgript. Gallai hyn effeithio ar ddilysrwydd y barnau o'u cymharu â marcio, ond un cwestiwn diddorol i ymchwilio iddo fyddai i ba raddau y mae'r agweddau hyn yn cael effaith yn ystod marcio confensiynol hefyd.

Roedd peth tystiolaeth y gallai cyfranogwyr fod wedi pwysoli rhai agweddau ar y cynllun marcio ychydig yn wahanol. Ystyriwyd yn helaeth fod hyn yn fantais gan BG, gan ei fod yn caniatáu i gyfranogwyr asesu a phwyso a mesur yn fwy priodol yr agweddau yr oeddent yn ystyried bod angen eu credydu. Mae hyn yn cyflwyno cwestiynau diddorol am sut y caiff y dull cyfannol ei ddefnyddio gan bob beirniad. Mae'n bosibl bod gwahanol agweddau'n dylanwadu ychydig yn wahanol ar feirniaid wrth wneud y penderfyniad sy'n creu amrywiad. Gellir creu amrywiad pellach gan feirniaid nad ydynt yn darllen y sgriptiau'n llawn, neu gan feirniaid sy'n dod ar draws sgriptiau mwy anarferol, megis sgriptiau sy'n hirach, neu ar bynciau llai cyfarwydd neu wahanol. Nid yw'n glir sut mae'r holl ystyriaethau hyn yn effeithio ar ddilysrwydd. Felly, mae angen ymchwil fwy cadarn i ymchwilio ymhellach i'r agwedd hon.

Hyd yn oed os yw BG mor ddilys â marcio, mae'n debyg y byddai amheuon dilysrwydd yn parhau o ystyried nad oes gan y broses benderfynu gyfannol y tryloywder sydd gan farcio dadansoddol confensiynol. Felly, wrth ddefnyddio BG, credwn y byddai angen arweiniad tebyg i'r cynllun marcio presennol ar aseswyr o hyd, er bydd angen llunio'r agweddau i'w hystyried i adlewyrchu natur fwy cyfannol y broses o wneud penderfyniadau. Efallai y bydd angen hyfforddiant arnynt hefyd i gynnal dealltwriaeth gyffredin o'r lluniad asesu, a chysondeb wrth ystyried gwahanol agweddau perthnasol ar ymateb.

Dangoswyd bod BG yn ddull dibynadwy, ond mae ei ddibynadwyedd yn dibynnu ar nifer y barnau a wnaed ar gyfer pob sgript. Fel y cyfryw, mae cysylltiad agos rhwng dibynadwyedd a dichonoldeb. Mae rhai astudiaethau wedi dangos na fyddai asesiad BG dibynadwy yn gofyn am fwy o amser na marcio traddodiadol (e.e. Kimbell, 2012; Steedle a Ferrara, 2016, ond gweler Wheadon et al., 2020b fel enghraifft o ganfyddiadau i'r gwrthwyneb).

Nid yw ein canfyddiadau'n diystyru'r posibilrwydd y bydd BG yn ddull ymarferol o asesu rhan (a) NEA Hanes a mathau tebyg o atebion estynedig, mae'n debyg. Fodd bynnag, byddai hyn yn dibynnu ar y lefelau dibynadwyedd a fyddai'n dderbyniol wrth ddefnyddio BG ar gyfer pwnc penodol neu hyd yn oed dasg asesu benodol (ac felly

nifer y barnau sydd eu hangen i gyrraedd y lefelau hynny). Hyd yn oed os cymerwn y mynegeion dibynadwyedd a gafwyd yn yr ymchwil hon ar eu golwg, ni wyddom sut y maent yn cymharu â dibynadwyedd marcio confensiynol gan nad oes gennym ddata ar ddibynadwyedd yr eilbeth.

Mae'r ymchwil hon yn awgrymu, er mwyn cynnal yr un faint o amser ar gyfer asesu, y byddai angen i athrawon gynnal rhwng 9 ac 11 o farnau fesul sgript. Byddai cynyddu nifer y barnau gyda'r nod o sicrhau mwy o ddibynadwyedd yn gwneud gweithredu BG ar gyfer yr ymatebion hyn yn llai ymarferol, oni bai bod y broses yn cael ei hoptimeiddio ymhellach. Er enghraifft, gallai athrawon neilltuo gradd dros dro i bob ymateb a fyddai'n cynyddu'r wybodaeth a geir yn dilyn y barnau, yn cynyddu dibynadwyedd BG, a hynny heb fod yn faich sylweddol. Felly, gellid ymchwilio ymhellach i'r opsiwn hwn os dymunir mwy o ddibynadwyedd.

Ar wahân i nifer y barnau sydd eu hangen, rhaid mynd i'r afael â sawl agwedd arall (allan o'r rhai a archwiliwyd yn yr astudiaeth hon) er mwyn gwneud BG yn ymarferol. Yn gyntaf, er mwyn lleihau'r amser sydd ei angen i brosesu ymatebion, dylai dysgwyr deipio eu hatebion. Yn ail, dylid capio hyd yr ymatebion i uchafswm o 1,500 o eiriau. Byddai hyn yn lleihau'r amser sydd ei angen i feirniadu, ond hefyd efallai'n gwella ansawdd y barnau. Yn olaf, oherwydd ystyriwyd bod beirniadu ymatebion ar wahanol bynciau yn fwy heriol ac yn llai dibynadwy ac effeithlon, yn ol pob tebyg, efallai y byddwn yn ystyried cyfyngu ar nifer y meysydd pwnc y mae un beirniad yn eu gweld i ddau, os nad hyd yn oed un. Fodd bynnag, dylid cynnal ymchwil bellach i archwilio sut mae pynciau lluosog yn effeithio ar ganlyniadau beirniadu a sut i fynd i'r afael â'r problemau posibl.

Mae cadwyno'r ymatebion, hynny yw, gweld un ymateb ddwywaith yn olynol, yn ffactor arall sy'n cyflymu'r beirniadu. Mae'r ddeddf a luniwyd gan Thurstone, a'r model ystadegol, yn rhagdybio bod barnau'n annibynnol, ac nad yw penderfyniad a wneir mewn un farn yn effeithio ar y penderfyniad mewn barn arall. Ni welsom dystiolaeth ystadegol gref fod yr ymateb wedi'i gadwyno yn cael ei ddewis yn amlach fel yr enillydd, ond ni allwn ddiystyru'r effaith. Mae angen ymchwilio ymhellach i'r mater hwn gan fod cadwyno'n bwysig o ran sicrhau bod BG yn ymarferol ar gyfer yr ymatebion ysgrifennu estynedig.

A pheth arall, gellid ystyried defnyddio BG fel offeryn asesu ffurfiannol ar gyfer ysgrifennu estynedig o lefelau cymhlethdod o'r fath. Fel y nodwyd uchod, soniodd rhai cyfranogwyr y buasent yn hoffi ei defnyddio ar gyfer adborth gan gymheiriaid yn eu dosbarthiadau. O'i defnyddio fel hyn, byddai BG yn rhoi adborth ar unwaith i bob dysgwr am ei berfformiad y gellir ei ddiweddarw drwy ddefnyddio BG dro ar ôl tro yn ystod pwnc. Yn ogystal, o'i defnyddio ar gyfer adborth gan gymheiriaid, dangoswyd bod BG yn gwella'n sylweddol gyflawniad dysgwyr a gafodd eu dosbarthu'n is i ddechrau (Seery et al., 2019). Os caiff ei defnyddio ar y lefel genedlaethol, gallai BG

gynnig gwybodaeth i athrawon am safonau ar draws ysgolion a hwyluso cyfnewid profiad proffesiynol. Mae hyn fel petai'n arbennig o ddefnyddiol os rhoddir mwy o ymreolaeth i ysgolion mewn perthynas â chynnwys y cwricwlwm. Mae'n werth sôn bod yr Adolygiad Annibynnol o'r Cwricwlwm a'r Trefniadau Asesu yng Nghymru gan yr Athro Donaldson (2015, t.97), a oedd yn sail i ddiwygio'r cwricwlwm ar hyn o bryd, yn sôn bod rhannu arfer da ymhlith ysgolion yn un o'r ffactorau ar gyfer gweithredu cwricwlwm da yn llwyddiannus. Byddai defnyddio BG ar gyfer asesu ffurfiannol hefyd yn haws o ystyried nad yw hyn yn wynebu'r un pwysau â'r asesiad crynodol lle mae llawer yn y fantol.

Un ffordd ddiddorol o ymchwilio ymhellach fyddai archwilio'r asesiad o ymatebion sydd â lefel uwch o gymhlethdod na'r rhai a ddefnyddir yn yr astudiaeth hon, megis ymatebion rhan (b) a Safon Uwch Hanes, neu ymatebion a roddwyd ar gyfer pynciau eraill.

Er mwyn ystyried BG yn opsiwn yn lle marcio confensiynol, mae angen gwneud rhagor o waith ymchwil i weld sut y gellid defnyddio'r dull hwn mewn cyd-destun asesu lle mae llawer yn y fantol ac ystyried agweddau sy'n mynd y tu hwnt i'r asesu gwirioneddol; er enghraifft, archwilio agweddau fel hawl dysgwyr i apelio, a'r agweddau sefydliadol ar weithredu BG.

Cyfeiriadau

- Barkaoui, K. (2011). Effects of marking method and rater experience on ESL essay scores and rater performance. *Assessment in Education: Principles, Policy & Practice*, 18(3), 279-293.
- Bramley, T. (2005). A rank-ordering method for equating tests by expert judgement. *Journal of Applied Measurement*, 6(2), 202–223.
- Bramley, T. (2007). Paired comparison methods. In P. E. Newton, J. Baird, H. Goldstein, H. Patrick & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards*. (pp. 246-294). London: Qualifications and Curriculum Authority.
- Bramley, T. (2015). *Investigating the reliability of adaptive comparative judgment*. Cambridge Assessment Research Report. Cambridge, 36.
- Bramley, T. and Vitello, S. (2019). The effect of adaptivity on the reliability coefficient in adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 26(1), 43-58.
- Brown, T. & Peterson, G. (2009). *An enquiry into the method of paired comparison: reliability, scaling, and Thurstone's law of comparative judgment*. Gen Tech. Rep. RMRS-GTR-216WWW. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station, 98 pages.
- Coertjens, L., Lesterhuis, M., De Winter, B. Y., Goossens, M., De Maeyer, S. & Michels, N. (2021). Improving Self-Reflection Assessment Practices: Comparative Judgment as an Alternative to Rubrics, *Teaching and Learning in Medicine*, 33(5), 525-535.
- Cromptvoets, E. A. V., Béguin, A. A., & Sijtsma, K. (2020). Adaptive pairwise comparison for educational measurement. *Journal of Educational and Behavioral Statistics*, 45, 316-338.
- Daal, T. V., Lesterhuis, M., Coertjens, L., Donche, V., & Maeyer, S. D. (2016). Validity of comparative judgement to assess academic writing: examining implications of its holistic character and building on a shared consensus. *Assessment in Education: Principles, Policy & Practice*, 26(1), 59-74.
- Donaldson, G. (2015). *Successful Futures: Independent Review of Curriculum and Assessment Arrangements in Wales*, WG23258.

- Gill, T. & Bramley, T. (2013). How accurate are examiners' holistic judgements of script quality? *Assessment in Education: Principles, Policy & Practice*, 20(3), 308-324.
- Hodge, K.J., & Morgan, G.B. (2017). Stability of infit and outfit compared to simulated estimates in applied setting. *Journal of Applied Measurement*, 18(4), 383-392.
- Holmes, S., Black, B., & Morin, C., (2018). *Marking reliability studies 2017: Rank ordering versus marking – Which is more reliable?* Ofqual publication. 33p.
- Jones, I. & Iglis, M. (2015). The problem of assessing problem solving: can comparative judgement help? *Educ Stud Math*, 89, 337–355.
- Jones, I., Swan, M. & Pollitt, A. (2015). Assessing mathematical problem solving using comparative judgement. *Int J of Sci and Math Educ*, 13, 151–177.
- Kellaghan, T. & Greaney, V. (2019). *The Reliability of Public Examinations, Public Examinations Examined*. Washington, DC: World Bank. © World Bank.
<https://openknowledge.worldbank.org/handle/10986/32352> License: CC BY 3.0 IGO.
- Kimbell, R. (2012). Evolving project e-scape for national assessment. *International Journal of Technology and Design Education*, 22, 135–155.
- Kimbell, R. (2021). Examining the reliability of adaptive comparative judgement (ACJ) as an assessment tool in educational settings. *International Journal of Technology and Design Education.*, 1-15.
- Marshall, N., Shaw, K., Hunter, J. & Jones, I. (2020). Assessment by comparative judgement: an application to Secondary Statistics and English in New Zealand. *NZ J Educ Stud*, 55, 49–71.
- McGrane, J. A., Humphry, S. M., & Heldsinger, S. A. (2018). Applying a Thurstonian, two-stage method in the standardized assessment of writing. *Applied Measurement in Education*, 31(4), 297–311.
- McMahon, S. & Jones, I. (2015) A comparative judgement approach to teacher assessment. *Assessment in Education: Principles, Policy & Practice*, 22(3), 368-389.
- Murphy S., & Yancey K.B., (2009). Construct and Consequence: Validity in Writing Assessment, in *Handbook of Research on Writing: History, Society, School, Individual, Text*, Routledge, p. 450-464.

Newhouse, C. P. (2014). Using digital representations of practical production work for summative assessment, *Assessment in Education: Principles, Policy & Practice*, 21:2, 205-220.

Pollitt, A. (2012b). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 19(3), 281-300.

Pollitt, A. & Crisp, V. (2004, September). *Could comparative judgements of script quality replace traditional marking and improve the validity of exam questions?* Paper presented at the British Educational Research Association Annual Conference, UMIST, Manchester.

Rangel-Smith, C. & Lynch, D. (2018, 18th – 21st June). *Addressing the issue of bias in the measurement of reliability in the method of adaptive comparative judgment.* Conference presentation at 36th International Pupils' Attitudes Towards Technology Conference Athlone Institute of Technology, Co. Westmeath, Ireland.

Rhead, S., Black, B., Pinot de Moira, A., (2018). *Marking Consistency Metrics.* Ofqual report, 48 pages.

Seery, N., Canty, D., & Phelan, P. (2012) The validity and value of peer assessment using adaptive comparative judgement in design driven practical education. *International Journal of Technology and Design Education*, (22), 205–226.

Steedle, J.T. & Ferrara, S. (2016). Evaluating comparative judgment as an approach to essay scoring, *Applied Measurement in Education*, 29(3), 211-223.

Tarricone, P. & Newhouse, C. (2016). Using comparative judgement and online technologies in the assessment and measurement of creative performance and capability, *International Journal of Educational Technology in Higher Education*, 13-16.

Thurstone, L.L. (1927). A Law of Comparative Judgment. *Psychology Review*, 34, 273-286 https://brocku.ca/MeadProject/Thurstone/Thurstone_1927f.html

Verhavert, S., Bouwer, R., Donche, V. & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement, *Assessment in Education: Principles, Policy & Practice*, 26(5), 541-562.

Wheadon, C., Barmby, P., Christodoulou, D. & Henderson, B. (2020 a). A comparative judgement approach to the large-scale assessment of primary writing in England, *Assessment in Education: Principles, Policy & Practice*, 27(1), 46-64.

Wheadon, C., de Moira, A. P., & Christodoulou, D. (2020 b). The classification accuracy and consistency of comparative judgement of writing compared to rubric-based teacher assessment. SocArXiv <https://doi.org/10.31235/osf.io/vzus4>

Whitehouse, C. & Pollitt, A. (2012). *Using adaptive comparative judgement to obtain a highly reliable rank order in summative assessment*, AQA Centre for Education Research and Policy, Manchester.

Whitehouse, C. (2013). *Testing the validity of judgements about geography essays using the adaptive comparative judgement method*, Centre for Education Research and Policy AQA Education.

